

**Der Vortesteffekt – wenn Tests bereits vor dem
Lernen unser Gedächtnis fördern**

Inaugural-Dissertation zur Erlangung der Doktorwürde
der Fakultät für Humanwissenschaften der Universität
Regensburg

vorgelegt von
JOHANNES BARTL
aus Landshut

Regensburg 2025

Erstgutachter (Betreuer): Prof. Dr. Karl-Heinz T. Bäuml

Zweitgutachter: Prof. Dr. Peter Fischer

Danksagung

Zuallererst gilt mein besonderer Dank Prof. Dr. Karl-Heinz Bäuml, der die Verfolgung einer akademischen Laufbahn für mich erst möglich gemacht hat. Durch seine fachliche Expertise, seine wertvollen Ratschläge und eine stets wertschätzende und offene Kommunikation fühlte ich mich jederzeit bestens unterstützt und betreut. Vor allem die für mich passende Balance aus vorgegebener Struktur und persönlicher Freiheit habe ich als sehr hilfreich und zugleich motivierend empfunden.

Zudem möchte ich mich ganz herzlich bei Oliver Kliegl bedanken, der das Projekt mit initiierte und betreute und mich während des gesamten Prozesses unterstützte. Das hat mir den Einstieg in das Projekt und die konsequente Verfolgung bis zum Abschluss sehr erleichtert, vielen Dank! Ein großer Dank gebührt auch Magdalena Abel, die mir bei der Mitarbeit an ihrem Projekt stets den Rücken für die Vollendung dieser Arbeit freihielt.

Des Weiteren gilt ein besonderer Dank meinen (aktuellen und ehemaligen) Kolleginnen und Kollegen Lukas Trißl, Verena Kriechbaum, Sarah Meixensperger, Marilena Hirsch, Anna Nickl, und Zsuzsanna Nemecz. Ich habe den persönlichen und fachlichen Austausch stets genossen und konnte wahnsinnig davon profitieren. Auch in schwierigen Situationen habe ich mich so nie allein gefühlt.

Für die vielfältige Unterstützung bei allen Belangen rund um das Thema formale Angelegenheiten möchte ich mich bei Petra Witzmann bedanken.

Ein aufrichtiges Dankeschön geht an Philipp Vogel. Wir haben uns im Studium gefunden und seither bist du mir immer ein verlässlicher Rückhalt, guter Zuhörer und vor allem ein bester Freund. Auch Teresa Schmidt möchte ich herzlich für deine freundschaftliche Unterstützung und den persönlichen und fachlichen Austausch danken.

Vor allem möchte ich mich bei meiner Schwester Anna bedanken, die mir während all der Zeit sowohl emotional als auch praktisch immer zur Seite stand und mich durch ihre vielfältigen Hilfen stets entlastete und unterstützte.

Zuletzt möchte ich mich noch bei meinen Eltern und Großeltern bedanken. Ich kann gar nicht in Worte fassen, wie dankbar ich bin, dass ihr in jeder Phase meines Lebens und meiner akademischen Laufbahn hinter mir steht und mich nach Kräften unterstützt und ich immer auf eure Unterstützung zählen kann. Vielen Dank, ohne euch wäre das nicht möglich gewesen.

Inhaltsverzeichnis

Zusammenfassung	6
1. Förderliche Effekte von Abrufübungen	7
2. Der Vortest-Effekt	12
2.1. Definition und Charakteristika	12
2.2. Befunde über die Robustheit dieses Effekts	16
2.3. Theoretische Erklärungsansätze	20
Elaborations-Hypothese	20
Aufmerksamkeits-Hypothese	21
Error-Correction-Hypothese	22
Kontext-Hypothese	23
Search-Set-Hypothese	24
Cue-Overload-Prinzip	26
3. Offene Fragen und Ziele der vorliegenden Arbeit	27
4. Der Einfluss wiederholter Tests auf die Erinnerungsleistung	30
4.1. Bisherige Befunde	30
4.2. Theoretische Erwartungen	33
4.3. Fragestellung der Experimente 1-4	36
5. Der Vortest-Effekt mit wiederholten Rateversuchen beim Test	39
5.1. Experiment 1	39
Methode	39
Ergebnisse	42
Diskussion	44
5.2. Experiment 2	44

Methode	44
Ergebnisse.....	46
Diskussion	47
5.3. Experiment 3	49
Methode	49
Ergebnisse.....	50
Diskussion	52
5.4. Experiment 4	53
Methode	53
Ergebnisse.....	55
Diskussion	58
Zusätzliche Analysen.....	59
5.5. Diskussion der Experimente 1-4	60
6. Die Rolle von semantischen Mediatoren für Test-Effekte.....	67
6.1. Bisherige Befunde	67
6.2. Theoretische Erwartungen.....	69
6.3. Fragestellung der Experimente 5-7	72
7. Die Rolle von Mediatoren für den Vortest-Effekt.....	76
7.1. Experiment 5	76
Methode	76
Ergebnisse.....	79
Diskussion	84
7.2. Experiment 6	85
Methode	85
Ergebnisse.....	86

Diskussion	88
7.3. Experiment 7	89
Methode	89
Ergebnisse.....	90
Diskussion	92
7.4. Diskussion Experimente 5-7.....	93
8. Der Einfluss geteilter Aufmerksamkeit auf Test-Effekte	98
8.1. Bisherige Befunde	98
8.2. Theoretische Erwartungen.....	101
8.3. Fragestellung der Experimente 8-9	105
9. Der Vortest-Effekt unter geteilter Aufmerksamkeit.....	108
9.1. Experimente 8a und 8b.....	108
Methode	108
Ergebnisse von Experiment 8a	111
Ergebnisse von Experiment 8b	114
Zusätzliche Analysen.....	116
Diskussion	117
9.2. Experimente 9a und 9b.....	119
Methode	119
Ergebnisse von Experiment 9a	120
Ergebnisse von Experiment 9b	122
Zusätzliche Analysen.....	125
Diskussion	126
9.3. Diskussion Experimente 8-9.....	127
10. Gesamtdiskussion	135

10.1. Zusammenfassung der Befunde	135
10.2. Theoretische Schlussfolgerungen.....	136
10.3. Einordnung der Befunde in die bisherige Literatur	143
10.4. Praktische Implikationen.....	145
10.5. Ansätze für zukünftige Forschung.....	148
10.6. Fazit	153
Literaturverzeichnis	155

Teile dieser Arbeit liegen vor als:

- Kliegl, O., Bartl, J., & Bäuml, K. H. T. (2024). Repeated guessing attempts during acquisition can promote subsequent recall performance. *Journal of Experimental Psychology: Applied*, 13(1), 63-70.
- Bartl, J., Kliegl, O., & Bäuml, K. H. T. (2024). The role of mediators for the pretesting effect. *Memory*, 32(3), 358-368.
- Bartl, J., Kliegl, O., & Bäuml, K. H. T. (2025). The pretesting effect under divided attention. *Psychological Research*, 89, 77.

Zusammenfassung

Das Bearbeiten eines Tests (Rauch - ?) bereits vor der Präsentation der Lerninformation (Rauch – *Nebel*) kann die Erinnerungsleistung für dieses Material im Vergleich zu Lernmaterial, das unmittelbar gelernt wird, verbessern. Dieser sogenannte Vortest-Effekt gilt als robustes Phänomen, das sich über unterschiedliche Arten von Lernmaterial, Testformate oder Stichproben hinweg finden lässt. Über die zugrundeliegenden kognitiven Mechanismen dieses Gedächtnisphänomens ist bisher jedoch nur wenig bekannt. In neun Experimenten wurden daher konkrete theoretische Annahmen der beiden prominentesten Erklärungsansätze - der Elaborations-Hypothese und der Aufmerksamkeits-Hypothese - überprüft. Zunächst wurde der Einfluss wiederholter Vortests auf den Vortest-Effekt untersucht (Experimente 1-4), um Erkenntnisse darüber zu gewinnen, ob es dabei zu einem Cue-Overload kommt, der die Erinnerungsleistung beeinträchtigt oder, ob mehrfache Abrufübungen zu einem zusätzlichen Vorteil führen, wie dies die beiden genannten Hypothesen vermuten würden. Mit den Experimenten 5 bis 7 sollte mit Hilfe semantischer Mediatoren die Annahme der Elaborations-Hypothese getestet werden, dass es während eines Vortests zu der Aktivierung eines semantischen Netzwerks und der Integration des Lernmaterials in dieses kommt. Die Vorhersage der Aufmerksamkeits-Hypothese, dass es durch Vortests zu einer gesteigerten Neugier und Motivation und damit zu einer aufmerksameren Enkodierung der Zielinformation kommt, wurden durch die Untersuchung des Effektes unter geteilter Aufmerksamkeit näher betrachtet (Experimente 8 und 9). Die Ergebnisse der Experimente 1 bis 4 zeigten dabei für unterschiedliches Lernmaterial eine Vergrößerung des Vortest-Effekts nach wiederholten Vortests. In den Experimenten 5 bis 7 war auch dann ein Vortest-Effekt nachzuweisen, wenn semantische Mediatoren als Hinweisreize für den Abruf der Zielinformation verwendet wurden. Die Experimente 8 und 9 offenbarten einen vergrößerten Vortest-Effekt für geteilte Aufmerksamkeit während des Vortests und einen ausbleibenden Vortest-Effekt für geteilte Aufmerksamkeit während des korrektiven Feedbacks. Diese Befunde legen nahe, dass sowohl Elaborationsprozesse als auch Aufmerksamkeitsressourcen eine wichtige Rolle für die Entstehung des Vortest-Effekts spielen. Die Auswirkungen auf weitere theoretische Erklärungsansätze dieses Effekts sowie die praktische Anwendung des Vortest-Paradigmas werden diskutiert.

1. Förderliche Effekte von Abrufübungen

„Unterlagen weg, wir schreiben jetzt einen Test!“: Sätze wie diese lassen den Puls von Schülerinnen und Schülern sowie von Studentinnen und Studenten schneller schlagen, denn Prüfungssituationen stellen häufig ein unliebsames Ereignis dar, das mit Druck und Stress einhergeht. So zeigte eine Befragung der Internationalen Hochschule im Jahr 2022 unter 1600 Personen unterschiedlicher Altersgruppen, dass 64.7 % der Deutschen schon einmal Prüfungsangst im schulischen oder universitären Kontext erlebt haben. Diese Prüfungsangst ist häufig verbunden mit körperlichen, emotionalen und kognitiven Symptomen. Je nach Art der Prüfung kann dies auch zu einem Anstieg des subjektiven und körperlichen Stresserlebens führen, wobei insbesondere Tests, die reproduktive Aufgaben beinhalten (d. h. Aufgaben in denen gelernte Inhalte aus dem Gedächtnis abgerufen werden sollen) zu einem stärkeren Anstieg des Cortisolspiegels führen können als das Bearbeiten von Problemlöseaufgaben (Minkley & Kirchner, 2012). Doch warum sind Prüfungen und insbesondere Gedächtnistests so gefürchtet? Im schulischen und universitären Kontext werden Tests vorrangig zur Erfassung der Lernleistung und zur Bestimmung des aktuellen Kenntnisstandes eingesetzt und gehen in vielen Fällen mit einer Benotung einher, was zu einer Empfindung von Leistungsdruck beitragen kann. Zudem können Lernende durch Tests auf eigene Unzulänglichkeiten oder Wissenslücken aufmerksam gemacht werden, was als affektiv unangenehm empfunden wird (z. B. Deci & Ryan, 2000). Es stellt sich daher die Frage, ob Tests und Prüfungen aus dem schulischen und universitären Kontext verschwinden sollen.

Diese Frage kann eindeutig mit nein beantwortet werden, denn insbesondere Gedächtnistests sind mehr als nur Instrumente zur Erfassung von Vorwissen. So kann der Abruf von Inhalten aus dem Gedächtnis diesen Inhalt auch verändern, indem dadurch bestehende Gedächtnisspuren modifiziert und neue Gedächtnisspuren und -verknüpfungen geschaffen werden (z. B. Rowland, 2014). Sogenannte Abrufübungen stellen also selbst ein Lernereignis dar und die Verwendung derartiger Tests kann daher als eine von vielen Lernstrategien bezeichnet werden. Da Menschen nicht nur im angewandten

Bildungskontext, sondern auch in alltäglichen Situationen mit dem Lernen und Abrufen von Gedächtnisinhalten konfrontiert werden (z. B. Im Urlaub versuchen wir uns Fakten über ein Gebäude oder Land einzuprägen, um diese nach dem Urlaub erzählen zu können), ist die Suche nach Möglichkeiten, gelernte Informationen bestmöglich behalten zu können, seit jeher von großem Interesse. Auf dieser Suche wenden Lernende eine Vielzahl von Lernstrategien an. Dazu zählen beispielsweise das Markieren oder Zusammenfassen von Inhalten, die Nutzung von Karteikarten, das Lernen in Lerngruppen oder das Anwenden des Gelernten in praktischen Beispielen und Transferaufgaben. Allerdings variiert die Häufigkeit der Anwendung bestimmter Lernstrategien in der Praxis. So untersuchten Karpicke et al. (2009) in ihrer Studie, welche Lernstrategien von Studierenden grundsätzlich genutzt und welche Lernstrategie bevorzugt eingesetzt werden. Dazu sollten Studierende in einem Fragebogen zunächst alle Lernstrategien auflisten, die von ihnen bereits angewandt wurden und diese anschließend nach der Häufigkeit ihrer Nutzung ordnen. Dabei wurde das wiederholte Lesen von Lerninhalten als häufigste Lernstrategie genannt (83.6 %), gefolgt von der Bearbeitung von Übungsaufgaben (42.9 %), der Nutzung von Karteikarten (40.1 %) und dem erneuten Abschreiben von Notizen (29.9 %). Abrufübungen in Form von Selbstabfragen hingegen wurden nur in 10.1 % der Fälle aufgelistet. Bei Betrachtung der bevorzugten Lernstrategie zeigte sich, dass das erneute Lernen mit 54.8 % am häufigsten als bevorzugte Strategie von den Studierenden genannt wurde. Das Bearbeiten von Übungsaufgaben sowie das Abschreiben von Notizen wurde mit jeweils 12.4 % am zweithäufigsten als favorisierte Option gewählt, während nur 1.1 % der Studierenden Abrufübungen als beliebteste Lernstrategie angaben (Karpicke et al., 2009).

Die Häufigkeit der Anwendung einer Lernstrategie gibt jedoch noch keine Auskunft darüber, wie gut mit dieser einzelnen Strategie tatsächlich gelernt werden kann. Es bleibt also die Frage, ob eine Lerntechnik zu Recht häufig zum Einsatz kommt oder, ob andere Methoden besser geeignet wären, um Lerninhalte langfristig behalten zu können. Um dieser Frage auf den Grund zu gehen, analysierten Dunlosky et al. (2013) in einem Überblicksartikel die Effektivität unterschiedlicher Lernstrategien. Die Autoren untersuchten dabei zehn ausgewählte Lernstrategien, darunter das erneute Lesen von Lerninhalten, das Durchführen einer Abrufübung, das Zusammenfassen oder Markieren von Informationen oder die Verwendung mentaler Bilder zur Visualisierung von Lerninhalten.

Diese Strategien wurden dabei hinsichtlich ihrer Wirksamkeit für unterschiedliche Lern-Bedingungen (z. B. Umfang des Lernens; Lesen vs. Zuhören), für unterschiedliche Stichprobeneigenschaften (z. B. Alter, Motivation, Arbeitsgedächtniskapazität), für unterschiedliche Materialien (z. B. Vokabeln, Texte) und für unterschiedliche Testformate (z. B. Cued-Recall, Problemlöseaufgaben, Rekognitionstests) betrachtet. Basierend auf diesen Kriterien und dem Vorliegen von Wirksamkeitsnachweisen der einzelnen Lernstrategien in einem angewandten Bildungskontext nahmen die Autoren eine abschließende Bewertung bezüglich des Nutzens und der Effektivität der jeweiligen Strategien vor. Diese wurde für das Zusammenfassen oder Markieren von Lerninhalten ebenso als gering eingeschätzt wie für das erneute Lesen von Informationen oder die Verwendung mentaler Bilder. Die Verwendung von Abrufübungen hingegen wurde als sehr nützlich und effektiv einsetzbar beurteilt (Dunlosky et al., 2013).

Abrufübungen liefern also nicht nur Hinweise über den aktuellen Wissensstand einer Person, sondern verbessern zusätzlich die Lernleistung für diese Information (Metcalf & Kornell, 2007). Ein Befund, der diesen positiven Effekt von Abrufübungen demonstriert, ist der sogenannte klassische Test-Effekt (für Überblick siehe Karpicke, 2017). Eine typische Prozedur zur Untersuchung des Test-Effekts besteht dabei aus einer Lernphase, in der den Versuchspersonen beispielsweise Cue-Target-Paare präsentiert werden (z. B. Stern – *Nacht*), gefolgt von einer Test-Phase. In dieser Test-Phase werden den Versuchspersonen die zuvor präsentierten Paare entweder erneut dargeboten oder die Probanden sollen nach der Präsentation des Cue-Wortes versuchen, das gelernte Target-Wort abzurufen (z. B. Stern - ?). Abschließend findet dann für die Versuchspersonen beider Bedingungen ein Test statt, bei dem alle gelernten Inhalte abgerufen werden sollen. Dabei zeigt sich typischerweise eine bessere Erinnerungsleistung für getestete als für wiederholt präsentierte Paare, was den klassischen Test-Effekt widerspiegelt (z. B. Rowland, 2014). Der Test-Effekt hat sich dabei als robustes und generalisierbares Phänomen erwiesen, das sich für unterschiedliche Lernmaterialien wie Wortpaare (z. B. Allen et al., 1969), Videos (z. B. Johnson & Mayer, 2009) oder Textpassagen (z. B. McDaniel et al., 2009; Wissman et al., 2011) zeigen ließ. Auch konnte dieser Effekt für verschiedene Altersgruppen (z. B. Aslan & Bäuml, 2016; Meyer & Logan, 2013; Rogalski et al., 2014) und Leistungsniveaus und sogar im angewandten Schulsetting mit realen Lerninhalten gefunden werden (Agarwal et al., 2021,

Roediger & Butler, 2011). Zudem stellte sich der positive Effekt von Tests auf die spätere Erinnerungsleistung als langlebig dar, da dieser nach mehreren Tagen, Wochen und sogar Monaten noch Bestand hatte (für Überblick siehe Yang et al., 2021).

Obwohl also das Durchführen einer Abrufübung eine erfolgreiche Lernstrategie darstellt und insbesondere im Vergleich mit der erneuten Präsentation von Inhalten zu einer besseren und längerfristigen Erinnerungsleistung für diese Inhalte führt, wird diese Strategie, wie zuvor beschrieben, deutlich seltener von Lernenden eingesetzt als das erneute Lernen (z. B. Roediger et al., 2009). Ein Grund dafür könnte sein, dass Lernenden dieser Vorteil von Tests nicht bewusst zu sein scheint. So werden die Versuchspersonen in einigen Studien zur Untersuchung des klassischen Test-Effekts gebeten, nach der Präsentation eines jeden Items oder im Anschluss an die gesamte Lernphase eine Einschätzung in Form sogenannter *Judgements of Learnings* (JOLs) darüber abzugeben, wie wahrscheinlich die gelernten Items in einem späteren Test korrekt abgerufen werden können. Dabei werden für Items der erneuten Lern-Bedingung typischerweise höhere JOLs abgegeben als für Items in der Test-Bedingung (z. B. Agarwal et al., 2008; Karpicke & Roediger, 2008), wohingegen die tatsächliche Erinnerungsleistung in einem finalen Test in der Test-Bedingung signifikant besser ausfällt als in der Lern-Bedingung. Es scheint also eine Diskrepanz zwischen der subjektiv angenommenen Lernleistung und dem tatsächlichen Lernerfolg bezüglich der Verwendung von Abrufübungen als Lernstrategie zu bestehen, die zu einem seltenen Einsatz dieser Technik im praktischen Lernalltag beitragen könnte. Zusammenfassend lässt sich also festhalten, dass die Literatur zum klassischen Test-Effekt zeigt, dass das Bearbeiten eines Tests nach dem Lernen die Erinnerungsleistung für diese Inhalte fördern kann.

Allerdings scheinen Tests nicht nur einen positiven Effekt auf das zuvor gelernte Material zu haben, sondern das Gedächtnis auch dahingehend zu beeinflussen, dass das nachfolgende Lernen von neuem Material ebenfalls gefördert wird, wie die Literatur zum sogenannten vorwärtsgerichteten Test-Effekt zeigt (z. B. Chan et al., 2018; Pastötter & Bäuml, 2014). In einer typischen Prozedur zur Untersuchung des vorwärtsgerichteten Test-Effekts werden den Versuchspersonen mehrere Listen mit Inhalten präsentiert, die gelernt werden sollen. Dabei findet im Anschluss an die Präsentation der einzelnen Listen für einen Teil der Versuchspersonen ein Test statt, bei dem die Inhalte der vorherigen Liste abgerufen werden sollen (Test-Bedingung), wohingegen der andere Teil der Versuchspersonen die

zuvor gelernte Liste erneut dargeboten bekommt (erneute Lern-Bedingung), bevor in beiden Bedingungen die nächste Liste präsentiert wird. Im Anschluss an die Lernphase der letzten Liste findet abschließend in beiden Bedingungen ein Test über den Inhalt der zuletzt gelernten Liste statt, häufig gefolgt von einem kumulativen Test über alle zuvor gelernten Listen. Die Ergebnisse zeigen dabei typischerweise eine bessere Erinnerungsleistung in der Test-Bedingung verglichen mit der erneuten Lern-Bedingung, sowohl im isolierten Test der zuletzt gelernten Liste als auch im kumulativen finalen Test, was dem vorwärtsgerichteten Test-Effekt entspricht. Zusätzlich führt die Bearbeitung von Tests zwischen den einzelnen Listen auch zu einer geringeren Anzahl an Interlist-Intrusionen im finalen Test, d. h. Items der anderen Lern-Listen werden in der Test-Bedingung seltener fälschlicherweise der letzten Liste zugeordnet als in der erneuten Lern-Bedingung (z. B. Szpunar et al., 2008). Die Durchführung einer Abrufübung führt also zu einer besseren Erinnerungsleistung und ebenso zu einer reduzierten Anzahl an Intrusionen und bewirkt damit einen Vorteil für das spätere Lernen neuer Informationen.

Insgesamt wirkt sich die Durchführung von Abrufübungen also sowohl auf das langfristige Behalten zuvor gelernter Informationen als auch auf das nachfolgende Lernen neuer Informationen positiv aus, was deren vielfältigen Einsatz als Lernstrategie ermöglicht. Allerdings haben die beiden zuvor beschriebenen Phänomene des klassischen Test-Effekts und des vorwärtsgerichteten Test-Effekts eine Gemeinsamkeit: Die Tatsache, dass den Versuchspersonen dabei zunächst Lern-Informationen präsentiert werden, bevor anschließend, nach dem Lernen, ein Test bearbeitet werden soll. Allerdings stellt sich hier die Frage, ob ein Test zwingend nach dem Lernen stattfinden muss, um einen vorteilhaften Effekt auf die Gedächtnisleistung haben zu können oder, ob es auch denkbar wäre, bereits vor der erstmaligen Präsentation von Lerninhalten einen Test durchzuführen.

2. Der Vortest-Effekt

2.1. Definition und Charakteristika

Tatsächlich ist die Bearbeitung von Tests nicht auf Zeitpunkte nach dem Lernen beschränkt. Diese können bereits vor der Präsentation des Lerninhalts stattfinden, weshalb diese Art der Abrufübung als sogenannter Vortest bezeichnet wird. Diese Vortests wurden und werden, ähnlich wie die klassischen Abrufübungen nach dem Lernen, überwiegend dazu verwendet, sowohl Lehrpersonen als auch den Lernenden selbst einen Überblick über den bisherigen Wissensstand bzw. das Vorwissen zu einem bestimmten Thema zu verschaffen (für Überblick siehe Black & Wiliam, 1998). So könnte eine Lehrperson beispielsweise zu Beginn einer Biologiestunde zum Thema Photosynthese die Einstiegsfrage „Wer weiß, welche Endprodukte es bei der Photosynthese gibt?“ stellen. Dieser Vortest könnte der Lehrperson folglich Aufschluss darüber geben, welche Lernenden bereits Vorwissen haben und so den Unterricht mit entsprechenden Antworten aktiv mitgestalten können. Zudem könnte der Vortest anzeigen, auf welche Lernenden besonders bei der Vermittlung der Lerninhalte geachtet und eingegangen werden sollte, da diese sich ohne Vorwissen möglicherweise schwerer beim Verstehen der nachfolgenden Inhalte tun könnten. Durch die Erkenntnisse eines Vortests können Lehrpersonen also folglich den Unterricht anpassen und personalisiertes Feedback erteilen (Bennett, 2011). Andererseits kann diese Art von Vortest auch den Lernenden selbst dabei helfen, individuelle Informationsdefizite aufzuzeigen, um den Fokus bei der nachfolgenden Unterrichtseinheit gezielt auf das Füllen dieser Lücken legen zu können. So könnte eine Person, die auf die Eingangsfrage „Welche Endprodukte der Photosynthese gibt es?“ bereits die Teilantwort „Sauerstoff“ geben kann, den Fokus in der nachfolgenden Unterrichtseinheit vor allem auf die Herleitung des zweiten Endproduktes „Glukose“ richten, um das bisherige Wissen zu ergänzen. Eine Person, die noch keines der beiden Endprodukte vorab benennen kann, würde möglicherweise einen stärkeren Fokus auf die Ausführungen zu beiden Endprodukten legen. Vortests wurden also vor allem zur formativen Bewertung eingesetzt, wobei in der Regel nicht berücksichtigt wurde, welche unmittelbaren Auswirkungen die Bearbeitung dieses Testes per se auf die

Erinnerungsleistung oder das Verständnis der Lerninhalte haben könnte (z. B. Heritage, 2007).

Im vorherigen Abschnitt konnte bereits veranschaulicht werden, dass das Durchführen einer Abrufübung nach dem Lernen nicht nur zur Erfassung der bisherigen Lernleistung beiträgt, sondern auch selbst ein Lernereignis darstellt, das sich positiv auf diese spätere Erinnerungsleistung des abgerufenen Materials auswirkt (klassischer Test-Effekt). Auch Vortests können nicht nur bei der Beurteilung von Vorwissen und der darauf aufbauenden Gestaltung von Unterrichtseinheiten oder der zielgerichteten Fokussierung beim Lernen des Unterrichtsinhaltes helfen, sondern sich ebenfalls positiv auf die spätere Lernleistung für diese Inhalte auswirken. Ein prototypischer Ablauf zur Untersuchung des Einflusses von Vortests auf die spätere Erinnerungsleistung, der die Grundlage für viele Studien zur Verwendung des Vortestparadigmas darstellt, stammt von Kornell et al. (2009). In mehreren Experimenten sollten Versuchspersonen dabei Wortpaare lernen, wobei die Versuchspersonen in einer (Baseline) Lern-Bedingung über die gesamte Dauer des Lern-Trials das vollständige Wortpaar, bestehend aus Cue-Wort und Target-Wort (z. B. Vogel – Feder), präsentiert bekamen, um dieses direkt zu lernen. Im Gegensatz dazu wurde den Versuchspersonen in der Vortest-Bedingung zu Beginn eines Lern-Trials zunächst nur das Cue-Wort präsentiert (z. B. Vogel - ?) und die Aufgabe der Versuchspersonen war es, das korrekte Target-Wort zu erraten, bevor im Anschluss die korrekte Lösung zu sehen war (z. B. Vogel – Feder). Dabei zeigte sich in einem finalen Cued-Recall-Test, in dem die Versuchspersonen die Cue-Wörter aller Wortpaare präsentiert bekamen und das dazugehörige Target-Wort abrufen sollten, eine signifikant bessere Erinnerungsleistung für Target-Wörter aus der Vortest-Bedingung verglichen mit der Lern-Bedingung, was den sogenannten Vortest-Effekt beschreibt. Dieses Befundmuster zeigte sich, obwohl die Präsentationsdauer eines Wortpaares in der Lern-Bedingung mit 10 s genau so lang war, wie die aufsummierte Dauer des Rateversuchs und des anschließend präsentierten Feedbacks in der Vortest-Bedingung, die jeweils 5s dauerten (Kornell et al., 2009). Zudem lässt sich in mehreren Studien auch bei der Betrachtung der Intrusionen, d. h. der Anzahl der im finalen Test fehlerhaften Antworten, eine signifikant niedrigere Anzahl an Intrusionen in der Vortest-Bedingung feststellen, was einen weiteren Vorteil von Vortests widerzuspiegeln scheint (z. B. Kliegl et al., 2024).

Ein positiver Einfluss von Vortests auf die spätere Erinnerungsleistung für diese Inhalte wurde bereits in den 1970er Jahren im Zuge der Forschung über die lernpraktischen Konsequenzen von Zusatzfragen, die vor, während oder nach dem Lernen beantwortet werden sollten, gezeigt (für Überblick siehe z. B. Rickards, 1979). So sollten beispielsweise Versuchspersonen in einer Studie von Rothkopf (1966) Ausschnitte über ein Buch zum Thema „Maritimes Leben“ lesen und sich die Inhalte der Ausschnitte für einen späteren Test einprägen, wobei ein Teil der Versuchspersonen einen Vortest mit einem Teil der späteren Testfragen bearbeiten sollte, ein anderer Teil der Versuchspersonen sollte die Inhalte direkt lernen. Die Erinnerungsleistung für Fragen, die vorgetestet wurden, war dabei signifikant besser als in der Bedingung, in der kein Vortest stattfand. Dieses Befundmuster konnte in darauffolgenden Studien teilweise repliziert werden (z. B. Hartley, 1973; Pressley et al., 1990), teilweise zeigten sich jedoch auch abweichende Befundmuster (z. B. Frase, 1968). Insgesamt muss jedoch festgehalten werden, dass die damalige Forschung zum Einfluss von Zusatzfragen auf die Lernleistung, anders als die „moderne“ Literatur zum Vortest-Effekt, häufig keine adäquate Kontrollgruppe oder zeitlich gematchte Alternativaufgaben beinhaltete oder den Effekt von Vortests und Abrufübung nach dem Lernen vermischte, weshalb die Interpretation des konkreten Einflusses von Vortests auf die spätere Erinnerungsleistung schwierig erscheint (Pan & Carpenter, 2023). Sowohl in dem früheren Forschungszweig als auch in der aktuellen Literatur zum Vortest-Effekt konnte aber gezeigt werden, dass das Bearbeiten eines Tests auch schon vor dem Lernen einen positiven Effekt auf die spätere Erinnerungsleistung haben kann.

Das mag zunächst kontraintuitiv erscheinen, denn anders als beim klassischen Test-Effekt-Paradigma, bei dem die Abrufübung nach dem Lernen stattfindet und der Test somit eine Abfrage bereits erworbenen Wissens darstellt, soll beim Vortest-Effekt-Paradigma im Test primär geraten werden, da die korrekte Lösung noch nicht bekannt ist. Dies hat zur Folge, dass der Anteil an korrekten Antworten in einem Vortest deutlich niedriger ausfällt als in einem Test, der nach dem Lernen stattfindet. So lag beispielsweise in der Studie von Halamish und Bjork (2011), in der das klassische Test-Effekt-Paradigma verwendet wurde, der Anteil korrekter Antworten im Test bei 82 %, wohingegen dieser Anteil in Studien zum Vortest-Paradigma häufig weniger als 5 % beträgt (z. B. 3.5 % in Kliegl et al., 2024; 4.4 % in Kornell et al., 2009). Dementsprechend untersucht das klassische Test-Effekt-Paradigma

vorrangig den Einfluss korrekter Testantworten auf die spätere Erinnerungsleistung, wohingegen das Vortest-Paradigma vor allem den positiven Effekt fehlerhafter Testantworten auf das spätere Erinnern betrachtet. Zusätzlich werden Items, die im Vortest korrekt erraten wurden, typischerweise von den Analysen des finalen Tests ausgeschlossen, womit bei der Verwendung von Vortests der isolierte Effekt von Fehlern auf das spätere Erinnern untersucht wird. Umso überraschender scheint daher zunächst der Befund, dass fehlerhafte Antworten in einem Test zu einem langfristig besseren Behalten der korrekten Informationen beitragen.

Lange Zeit war die vorherrschende Meinung, dass erfolglose Abruftests negative Konsequenzen auf das spätere Erinnern haben, was sich insbesondere in der Literatur zum *errorless learning* abbildet. Dieser Forschungszweig vertritt die Ansicht, dass Lernen insbesondere dann effektiv ist, wenn Fehler während des Lernens minimiert werden (z. B. Skinner, 1958) und hat eine lange und einflussreiche Tradition in der Psychologie (Guthrie, 1952). Dieser Ansatz beruht unter anderem auf der theoretischen Annahme, dass das Auftreten von Fehlern während des Lernprozesses die Unterscheidbarkeit zwischen den fehlerhaften Items und der korrekten Lösung reduziert. Dadurch steigt die Wahrscheinlichkeit, dass diese Items beim Abruf in Konkurrenz zur korrekten Antwort treten und somit den Zugang zur richtigen Lösung erschweren (z. B. Postman & Underwood, 1973). Zudem geht dieser Forschungszweig davon aus, dass das Bearbeiten eines Tests einen Abrufprozess darstellt, der die Gedächtnisspur zu der entsprechenden gegebenen Antwort stärkt. Wird nun im Lernprozess die korrekte Antwort genannt, so sollte diese die Abrufroute zu der korrekten Lösung stärken, wohingegen das Produzieren eines Fehlers während eines Tests die Gedächtnisspur zu diesem Fehler stärken und damit in einem späteren finalen Test zu einer Wiedererinnerung an diesen Fehler führen sollte (z. B. Bjork, 1975). Dazu passt beispielsweise der Befund, dass, wenn Lernende in einem Multiple-Choice-Test einen Fehler machen, dieser Fehler auch in einem finalen Test mit höherer Wahrscheinlichkeit erneut auftritt (Marsh et al., 2007). Obwohl viele Experimente zur Überprüfung dieses Ansatzes aus Studien zum Lernverhalten von Tieren stammen, haben diese auch die Präferenz für die angewandten Lernmethoden im pädagogischen Setting mitbeeinflusst.

Auch weiterhin neigen Lehrpersonen ebenso wie Lernende dazu, Fehler während des Lernens zu vermeiden, da diese Strategie als weniger erfolgsversprechend eingeschätzt wird

(z. B. Pan et al., 2020; Yang et al., 2017). In der Studie von Pan et al. (2020) gaben die Versuchspersonen beispielsweise an, Fehler gezielt vermeiden zu wollen, selbst, wenn ihnen bewusst war, dass Fehler einen natürlichen Bestandteil des Lernprozesses darstellen. In einer weiteren Studie äußerten Lernende ebenfalls eine starke Präferenz dafür, sich Lerninhalte durch Lesen anzueignen, anstatt einen Test zu absolvieren, insbesondere dann, wenn ihnen die Lerninhalte noch unbekannt oder wenig vertraut waren (Kirk-Johnson et al., 2019). Huelser und Metcalfe (2012) baten in ihrer Studie die Versuchspersonen, die Effektivität zweier Lernmethoden auf die spätere Erinnerungsleistung einzuschätzen: zum einen die einer Lern-Bedingung, in der ohne Fehler gelernt wird, zum anderen die einer Vortest-Bedingung, in der vor dem Lernen häufig Fehler produziert werden. Dabei wurde die fehlerlose Lern-Bedingung als signifikant erfolgreicher eingeschätzt. Die Versuchspersonen glaubten also, dass das Produzieren von Fehlern zu einer geringeren Erinnerungsleistung führt. Tatsächlich ergab sich bei Betrachtung der Erinnerungsleistung im finalen Test aber eine bessere Erinnerungsleistung für die Vortest-Bedingung verglichen mit der Lern-Bedingung (Huelser & Metcalfe, 2012). Entgegen der Erwartung von Lernenden und Lehrpersonen scheint also das Produzieren von Fehlern während des Lernens nicht nur keinen negativen, sondern im Gegenteil sogar einen positiven Effekt auf die spätere Erinnerungsleistung zu haben. Dies zeigt sich vor allem im Phänomen des Vortest-Effekts.

2.2. Befunde über die Robustheit dieses Effekts

Wie zuvor beschrieben, liegt der Ursprung der Literatur über den Vortest-Effekt bereits einige Jahrzehnte zurück und die „moderne“ Forschung über diesen Effekt fand ihren Startpunkt vor allem in den 2009 veröffentlichten Studien von Kornell et al., sowie Richland et al. Seither wurde der Vortest-Effekt in einer Vielzahl von Studien untersucht, wobei unterschiedliches Lernmaterial, Versuchspersonen verschiedener Altersstufen, vielfältige Testformate während des Vortests und des finalen Tests, variierende Retentionsintervalle und unterschiedliche Anwendungsbereiche verwendet wurden, um die Generalisierbarkeit der Befunde zu überprüfen.

Am häufigsten wurden in der bisherigen Literatur zur Untersuchung des Vortest-Effekts Wortpaare als Lernmaterial verwendet. Diese Wortpaare bestehen dabei typischerweise aus einem Cue-Wort (z. B. Haus) und einem Target-Wort (z. B. Garage).

Dabei führt das Bearbeiten eines Vortests, bei dem die Versuchspersonen das Target-Wort zu einem gegebenen Cue-Wort erraten sollen (z. B. Haus - ?), in einem späteren finalen Test zu einer besseren Erinnerungsleistung für das Target-Wort (z. B. Garage) als das bloße Lernen des intakten Wortpaares (z. B. Bridger & Mecklinger, 2014; Kornell et al., 2009; Slamecka & Fevreiski, 1983). Dieses Befundmuster ließ sich in zahlreichen Studien replizieren, vor allem dann, wenn zwischen dem Cue-Wort und dem Target-Wort ein semantischer Zusammenhang besteht (für Überblick siehe Pan & Carpenter, 2023). Weniger eindeutig ist allerdings die Befundlage für die Verwendung von semantisch unverwandten Wortpaaren (z. B. Bauch - Gras), wobei in einigen Studien der Erinnerungsvorteil für vorgetestete Wortpaare in einem finalen Test ausblieb (z. B. Grimaldi & Karpicke, 2012; Knight et al., 2012). Andere Studien hingegen konnten den Vortest-Effekt auch für dieses Material finden (z. B. Potts & Shanks, 2014).

Neben Wortpaaren konnte der Vortest-Effekt bereits auch für die Verwendung von Vokabeln (Potts & Shanks, 2014) und Fakten (Kornell et al., 2009) demonstriert werden. So sollten die Versuchspersonen in Experiment 1 der Studie von Kornell et al. (2009) eine Reihe fiktiver und nicht-fiktiver Fakten lernen, die als Frage mit der dazugehörigen Antwort präsentiert wurden (z. B. „Wer schoss im 14. Jahrhundert mit einem Pfeil einen Apfel vom Kopf seines Sohnes? - Wilhelm Tell“). Für die Hälfte der Fakten wurde dabei die Frage direkt mit der korrekten Antwort präsentiert, während für die andere Hälfte der Fakten zunächst nur die Frage zu sehen war und die Versuchspersonen zuerst die dazugehörige Antwort erraten sollten, bevor diese präsentiert wurde. Fakten, für die ein Vortest bearbeitet werden sollte, wurden hier in einem finalen Test besser erinnert als Fakten, die direkt gelernt werden konnten (Kornell et al., 2009; siehe auch Kornell, 2014; Vaughn et al., 2017).

Außerdem ließ sich der Vortest-Effekt auch für Prosatexte zu unterschiedlichen Themenbereichen (z. B. Hausman & Rhodes, 2018; Sana & Carpenter, 2023) und für Videosequenzen (z. B. Carpenter & Toftness, 2017; James & Storm, 2019; Toftness et al., 2018) finden. Eine aktuelle Studie von Willis et al. (2020) liefert zudem erste Hinweise darauf, dass sich die Vorteile von Vortests möglicherweise sogar auf das Lernen angewandter praktischer Fähigkeiten übertragen lassen. In dieser Studie sollten Medizin-Studenten das Verfahren der Laparoskopie erlernen, wobei den Studierenden entweder direkt ein Instruktionsvideo präsentiert wurde oder die Studierenden das Verfahren zunächst in

einer Simulation ausprobieren sollten, bevor das Instruktionsvideo zu sehen war. In einem finalen Test gelang es den Studierenden in der Vortest-Bedingung die Laparoskopie schneller und mit weniger Fehlern durchzuführen (Willis et al., 2020).

Wie in der gerade eben beschriebenen Studie, wurde eine Vielzahl der bisherigen Studien zur Untersuchung des Vortest-Effekts mit einer jungen erwachsenen und überwiegend studentischen Stichprobe durchgeführt. Bisher konnte ein Vorteil von Vortests auf die spätere Erinnerungsleistung aber auch schon für fünfjährige Kinder (Carneiro et al., 2018), für Schüler der vierten und fünften Klasse (de Lima & Jaeger, 2020) sowie Schulkinder im Alter zwischen 11 und 13 Jahren (Metcalf & Finn, 2012) und für ältere gesunde Erwachsene (Cyr & Anderson, 2012, 2015) gezeigt werden. In den Studien von Cyr und Anderson (2012, 2015) lag das durchschnittliche Alter der älteren Versuchspersonen zwischen 70 und 72 Jahren, und die Versuchspersonen sollten Wörter aus unterschiedlichen semantischen Kategorien lernen. Diese Wörter wurden entweder direkt gelernt (z. B. Nutztier – Schwein) oder es sollte vor dem Lernen ein Rateversuch bei vorgegebener Kategorie generiert werden (z. B. „Nennen Sie ein Nutztier“). In einem späteren finalen Rekognitionstest war die Erinnerungsleistung für Wörter in der Vortest-Bedingung signifikant besser als in der Lern-Bedingung (Cyr & Anderson, 2012).

Für die Generalisierbarkeit des Vortest-Effekts spielt es jedoch eine wichtige Rolle, ob sich dieser Effekt auch unter Verwendung verschiedener Testformate finden lässt. Am häufigsten wurde in bisherigen Studien zur Untersuchung des Vorteils von Vortests das Cued-Recall-Testformat verwendet. Dabei wird den Versuchspersonen sowohl im Vortest als auch im finalen Test ein Cue-Wort präsentiert (z. B. Palme) und die Versuchspersonen sollen das dazugehörige Target-Wort erraten bzw. abrufen (z. B. Urlaub), wobei typischerweise die Target-Wörter der Vortest-Bedingung besser erinnert werden als die Target-Wörter der Lern-Bedingung (z. B. Kornell et al., 2009). In der zuvor beschriebenen Studie von Cyr und Anderson (2012) hingegen wurde ein Rekognitionstest verwendet, bei dem die präsentierten Items als „alt“ oder „neu“ eingeschätzt werden sollten, wobei sich ebenfalls ein Erinnerungsvorteil in der Vortest-Bedingung finden ließ. Dieser Befund wurde auch in weiteren Studien unter Verwendung eines Rekognitionstests repliziert (z. B. Seabrooke et al., 2019). Zudem konnte der Vortest-Effekt auch unter Verwendung eines

Multiple-Choice-Testformats (z. B. Potts & Shanks, 2014) und mittels eines freien Abrufes (z. B. Clark, 2016) gezeigt werden.

Von hoher Relevanz ist auch die Frage, wie lange dieser Erinnerungsvorteil in der Vortest-Bedingung anhält. Um diese Frage zu beantworten, hilft die Betrachtung des Effekts über verschiedene Retentionsintervalle, d. h. über unterschiedlich lange Zeiträume, die zwischen dem Lernen des Materials und dem Abruf dieses Materials in einem finalen Test vergehen. Bisherige Studien zum Vortest-Effekt haben gezeigt, dass dieser Effekt unmittelbar nach dem Lernen (z. B. Kornell, 2014; Overoye et al., 2021), 30 Minuten nach dem Lernen (Kliegl et al., 2024) sowie noch nach einem Tag (z. B. Kornell et al., 2009), nach zwei Tagen (Yan et al., 2014) und sogar bis zu einer Woche nach dem Lernen (z. B. Kliegl et al., 2024; Richland et al., 2009) gefunden werden kann.

Während bei vielen der zuvor beschriebenen Studien die Datenerhebung im Labor stattfand (z. B. Kornell et al., 2009; Vaughn et al., 2017), konnte der Vortest-Effekt in neueren Untersuchungen auch dann nachgewiesen werden, wenn die Datenerhebung online via Zoom (Kliegl et al., 2023, 2024) oder in einem angewandten Lernsetting stattfand (Carpenter et al., 2018; Soderstrom & Bjork, 2023). So wurde die Studie von Soderstrom und Bjork (2023) in einer Vorlesung über psychologische Forschungsmethoden durchgeführt, wobei die Studierenden der Universität zu Beginn einer jeden Vorlesungssitzung einen Multiple-Choice-Test mit Fragen über die Hälfte der Themen der nachfolgenden Vorlesungseinheit präsentiert bekamen, wohingegen für die andere Hälfte der Themen kein Vortest stattfand und diese Inhalte nur gelernt wurden. In einem späteren finalen Test wurde das Material der vorgetesteten Themenbereiche besser erinnert als das direkt gelernte Material (Soderstrom & Bjork, 2023).

Zusammenfassend lässt sich festhalten, dass es sich bei dem Vortest-Effekt um ein robustes Phänomen zu handeln scheint, dass sich vielfältig replizieren ließ. Allerdings ist noch unklar, durch welche kognitiven Mechanismen dieser Effekt zu Stande kommt. Deshalb sollen nachfolgend unterschiedliche Erklärungsansätze des Vortest-Effekts beschrieben werden.

2.3. Theoretische Erklärungsansätze

Elaborations-Hypothese

Die Elaborations-Hypothese stellt einen der bekanntesten Erklärungsansätze für den Befund des klassischen Test-Effekts dar (Carpenter, 2009, 2011), und eine Variante der Elaborations-Hypothese wird häufig auch zur Erklärung des Vortest-Effekts verwendet (z. B. Endres et al., 2017, Huelser & Metcalfe, 2012). Die Elaborations-Hypothese des Vortest-Effekts nimmt dabei an, dass die Präsentation eines Cue-Wortes (z. B. Teller - ?) während eines initialen Vortests, gepaart mit der Aufforderung an die Versuchspersonen, das dazugehörige Target-Wort zu erraten, zu einer Aktivierung von Gedächtnisrepräsentationen führt, die mit dem Cue-Wort verwandt sind (z. B. *Essen, Messer, Tisch*). Wenn den Versuchspersonen nun direkt im Anschluss korrekatives Feedback in Form des vollständigen Cue-Target-Paares (z. B. Teller - *Gabel*) präsentiert wird, sollen die voraktivierten Gedächtnisrepräsentationen mit dem Cue-Target-Paar verknüpft werden, was zu einer detaillierteren Gedächtnisspur für diese Inhalte führt. Dieses Konzept basiert auf dem Ansatz der sogenannten *spreading activation* (z. B. Collins & Quillian, 1972), der davon ausgeht, dass die Aktivierung bei einzelnen „Knoten“ im semantischen Netzwerk beginnt (in diesem Fall bei dem Cue-Wort) und sich anschließend auf benachbarte, eng verknüpfte „Knoten“ ausbreitet (in diesem Fall die verwandten Gedächtnisrepräsentationen) bis am Ende der „Knoten“ der Target-Wörter erreicht wird. In Trials der Lern-Bedingung (ohne Vortest) sollte es aufgrund der passiveren Verarbeitung der Information kaum zu einer Aktivierung verwandter Gedächtnisrepräsentationen kommen. Findet am Ende des Experimentes ein finaler Test statt, bei dem den Versuchspersonen erneut das Cue-Wort präsentiert wird und das dazugehörige Target-Wort abgerufen werden soll, könnte dies zu einer Aktivierung der verknüpften „Knoten“ und somit zu einer Reaktivierung der Information führen, die während des Vortests mit dem Cue-Target-Paar verknüpft wurde. Diese reaktivierten Informationen sollten dann als semantische Mediatoren fungieren und dadurch den Abruf des korrekten Target-Wortes vereinfachen. Der Vortest-Effekt sollte dementsprechend durch die unterschiedliche Verfügbarkeit dieser für den Abruf des Target-Wortes hilfreichen Informationen in der Vortest-Bedingung und der Lern-Bedingung zustande kommen.

Die Annahmen dieses theoretischen Erklärungsansatzes konnten bisher durch zwei zentrale Befunde gestützt werden. So konnte in einer Studie von Huelser und Metcalfe (2012) gezeigt werden, dass Vortests nicht nur zu einer besseren Erinnerungsleistung für die Target-Items führen, sondern auch den Abruf dieser Items verlangsamen. Dies könnte gemäß der Elaborations-Hypothese daran liegen, dass der Versuch, das Target-Wort eines zuvor vorgetesteten Wortpaares abzurufen, zunächst zu der Reaktivierung der semantisch verknüpften Information führt, durch die wiederum letztlich der Zugang zu dem Target-Wort hergestellt wird. Dieser zusätzliche Schritt im Abrufprozess sollte mehr Zeit erfordern und zu einer langsameren Antwort führen. Zudem finden sich mehrere Studien, in denen kein Vortest-Effekt gezeigt werden konnte, wenn das Lernmaterial aus unverwandten Wortpaaren (z. B. Tier - *Rauch*) bestand (z. B. Grimaldi & Karpicke, 2012). Dieser Befund passt gut zu den Annahmen der Elaborations-Hypothese, da bei dieser Art von Material das Generieren eines Rateversuchs zu einer Aktivierung von Information führen könnte, die mit dem Cue-Wort verwandt ist. Diese Information sollte aufgrund einer fehlenden semantischen Verknüpfung mit dem Target-Wort in einem späteren finalen Test aber nicht als semantischer Mediator fungieren und den Abruf des Target-Wortes erleichtern können, weshalb sich in der Vortest-Bedingung verglichen mit der Lern-Bedingung kein Erinnerungsvorteil mehr für die Target-Wörter zeigen sollte.

Aufmerksamkeits-Hypothese

Da es jedoch auch Studien gibt, die einen Vortest-Effekt auch für semantisch nicht verwandtes Material zeigen konnten (z. B. Potts & Shanks, 2014), werden abgesehen von Elaborationsprozessen noch weitere kognitive Mechanismen für die Erklärung des Vortest-Effekts herangezogen. Hierbei spielen vor allem Aufmerksamkeitsprozesse eine wichtige Rolle, weshalb die Aufmerksamkeits-Hypothese, neben der Elaborations-Hypothese, die prominenteste Erklärung für den Vortest-Effekt darstellt. Diese Aufmerksamkeits-Hypothese geht davon aus, dass das Generieren eines Fehlers während der Bearbeitung eines Vortests den Versuchspersonen ein Informationsdefizit zwischen ihrem aktuellen Wissenstand und der für die korrekte Beantwortung des Vortests erforderlichen Information bewusst macht. Durch die Präsentation einer Frage oder eines Cue-Wortes während des Vortests und die Aufforderung, die korrekte Lösung anzugeben, könnte den Versuchspersonen verdeutlicht werden, dass sie die korrekte Antwort noch nicht kennen und

sie deshalb raten müssen. Das Generieren einer eigenen Antwort wiederum sollte gemäß der Aufmerksamkeits-Hypothese die Neugierde der Versuchspersonen wecken und die Motivation steigern, die eigene Wissenslücke füllen zu wollen. Um diese Wissenslücke zu schließen, werden bei der anschließenden Präsentation des korrektiven Feedbacks am Ende des Vortest-Trials mehr Aufmerksamkeitsressourcen für die Verarbeitung der korrekten Lösung aufgewendet. Diese intensivere Auseinandersetzung mit der Zielinformation sollte zu einer effektiveren Enkodierung dieser beitragen. Im Gegensatz dazu entfällt dieser Prozess in der Lern-Bedingung, da hier direkt die korrekte Lösung präsentiert wird und gelernt werden kann. Dadurch fehlt hier sowohl die bewusste Wahrnehmung einer Wissenslücke als auch die durch Neugierde gesteigerte Aufmerksamkeit bei der Enkodierung der Zielinformation. Der Vortest-Effekt resultiert somit gemäß der Aufmerksamkeits-Hypothese aus der aufmerksameren Enkodierung der Zielinformation in der Vortest-Bedingung, die in einem finalen Test in einen verbesserten Abruf der Information mündet und sich als Erinnerungsvorteil für die Vortest-Bedingung äußert (Potts & Shanks, 2014; Zawadzka & Hanczakowski, 2019).

Ein Befund, der die Annahmen der Aufmerksamkeits-Hypothese stützt, stammt aus einer Studie von Potts et al. (2019). In dieser Studie sollten die Versuchspersonen ihre subjektive Neugier, die korrekte Antwort lernen zu wollen, entweder vor dem Generieren eines eigenen Rateversuchs, nach dem Generieren eines Rateversuchs oder gänzlich ohne das Generieren eines Rateversuchs (Lern-Bedingung) angeben. Dabei wurden höhere Ratings nach dem Generieren eines Rateversuchs abgegeben, was darauf hindeutet, dass Vortests tatsächlich die Neugier der Versuchspersonen auf die korrekte Antwort erhöhen können, was sich wiederum in einer aufmerksameren Enkodierung der entsprechenden Information äußern sollte (Potts et al., 2019).

Error-Correction-Hypothese

Ein weiterer Erklärungsansatz, der eine andere Perspektive darauf einnimmt, inwiefern das Produzieren von fehlerhaften Antworten während eines Vortests zur Entstehung des Vortest-Effekts beiträgt, ist die Error-Correction-Hypothese (Carrier & Pashler, 1992). Dieser Theorie zufolge führt die Kombination aus dem Generieren eines fehlerhaften Rateversuchs und der Präsentation der korrekten Antwort durch das unmittelbar nachfolgende korrektive Feedback zur Entstehung eines Fehlersignals. Dieses Fehlersignal

führt zu einer Aktivierung von Fehlerkorrekturmechanismen, die zu einer Unterdrückung der Verknüpfung zwischen dem Cue-Wort und der falschen Antwort beitragen und somit eine Hemmung der inadäquaten Abrufroute zwischen dem Cue-Wort und dem Rateversuch bedingen. Gleichzeitig erhöht dies die relative Assoziationsstärke zwischen dem Cue-Wort und dem Target-Wort, d. h. die korrekte Abrufroute wird gestärkt (Knight et al., 2012). Infolgedessen steigt die Wahrscheinlichkeit, dass bei der Präsentation des Cue-Wortes in einem finalen Test die richtige Abrufroute aktiviert und der Abruf des Target-Wortes erleichtert wird, wodurch der Vortest-Effekt zustande kommt. Die Error-Correction-Hypothese geht folglich davon aus, dass fehlerhafte Antworten gezielt gehemmt werden, um den Zugang zu dem korrekten Cue-Target-Paar zu erleichtern, was im Einklang mit konnektionistischen Modellen des Lernens durch Fehlerkorrekturmechanismen steht (McClelland et al., 1986). Dieser Erklärungsansatz passt gut zu Studien, die auch für semantisch unverwandtes Material einen Vortest-Effekt zeigen konnten (z. B. Potts & Shanks, 2014). Denn gemäß der Error-Correction-Hypothese sollte es auch ohne vorbestehende semantische Verknüpfung zwischen dem Lernmaterial zu einem Fehlersignal und den nachfolgend ausgelösten Fehlerkorrekturmechanismen kommen, die den vorteilhaften Effekt von Fehlern beim Lernen verursachen (Mera et al., 2022).

Kontext-Hypothese

Des Weiteren lässt sich auch die episodische Kontext-Hypothese (z. B. Metcalfe & Huelser, 2020; Overman et al., 2021) zur Erklärung des Vortest-Effekts heranziehen. Dieser theoretische Ansatz basiert auf der Annahme, dass während des Enkodierens von Lerninformation (z. B. Cue-Target-Paare) nicht nur die Information selbst, sondern auch Merkmale des zeitlichen Kontexts, der zum Zeitpunkt des Lernens vorhanden ist, mit abgespeichert werden (Estes, 1955; Mensink & Raaijmakers, 1988). Da angenommen wird, dass der Abruf von Inhalten aus dem Gedächtnis zu einem schnelleren Wechsel des Kontexts führt (Jang & Huber, 2008), könnten Lerninhalte in der Vortest-Bedingung folglich mit einem detaillierten zeitlichen Kontext, bestehend aus einer integrierten Repräsentation sowohl des Vortest-Kontexts als auch des Lern-Kontexts, verknüpft sein. Lerninhalte der Lern-Bedingung hingegen sollten ausschließlich mit dem Lern-Kontext verknüpft sein (Hunt et al., 2011). Zusätzlich könnte es in der Vortest-Bedingung auch zu einer stärkeren Verknüpfung des Lerninhaltes mit dem mentalen Kontext kommen als in der Lern-

Bedingung (z. B. Hopper & Huber, 2018). Bei einem späteren finalen Test könnte die erneute Präsentation des bereits vorgetesteten Materials zu einer Reaktivierung des Kontexts aus der Lern-Phase führen. Dieser reaktivierte Kontext sollte in der Vortest-Bedingung aufgrund der Verfügbarkeit detaillierterer zeitlicher und mentaler Kontextmerkmale den Zugang zu der korrekten Zielinformation erleichtern und so den Erinnerungsvorteil in der Vortest-Bedingung (verglichen mit der Lern-Bedingung) ermöglichen.

Ein Befund, der im Einklang mit der Kontext-Hypothese des Vortest-Effekts steht, stammt von Metcalfe und Huelser (2020), die in ihrer Studie Wort-Triplets bestehend aus einem Target-Wort (Hand) und dazu semantisch kongruenten (Handgelenk, Handfläche) oder inkongruenten Cue-Paaren (Baum, Palme) verwendeten. Die Versuchspersonen sollten dabei in der Vortest-Bedingung einen Rateversuch für ein drittes zu dem Cue-Paar passendes Wort generieren, während in der Lern-Bedingung das vollständige Wort-Triplet präsentiert wurde. In einem späteren finalen Cued-Recall-Test sollten die Versuchspersonen das korrekte Target-Wort sowie den eigenen Rateversuch abrufen, wobei sich nur dann ein Erinnerungsvorteil für Target-Wörter in der Vortest-Bedingung verglichen mit der Lern-Bedingung zeigte, wenn sich die Versuchspersonen auch an den eigenen Rateversuch erinnern konnten. Dies könnte darauf hindeuten, dass die Erinnerung an den eigenen Rateversuch eine Reaktivierung des episodischen Kontexts der Lern-Phase widerspiegelt, die wiederum den Zugang zu dem korrekten Target-Wort erleichtert. Findet diese Kontextreaktivierung hingegen nicht statt, was sich in einer ausbleibenden Erinnerung an den eigenen Rateversuch äußern könnte, gibt es gemäß der Kontext-Hypothese keinen Grund mehr, warum der Abruf des Target-Wortes in der Vortest-Bedingung leichter fallen sollte als in der Lern-Bedingung. Hier sollte gemäß diesem Erklärungsansatz also der Vortest-Effekt ausbleiben (Metcalfe & Huelser, 2020).

Search-Set-Hypothese

Abschließend soll noch die Search-Set-Hypothese zur Erklärung des Vortest-Effekts beschrieben werden (Grimaldi & Karpicke, 2012). Ähnlich wie die Elaborations-Hypothese geht dieser Erklärungsansatz davon aus, dass es bei der Bearbeitung des Vortests im Zuge des Generierens eines eigenen Rateversuchs zu einem Suchprozess nach der korrekten Lösung kommt, der zu der Aktivierung vieler semantisch mit dem Cue-Wort verwandter Items führt. Das sogenannte Search-Set beinhaltet dabei neben unterschiedlicher verwandter

Information auch die korrekte Lösung. So könnte beispielsweise die Präsentation des Cue-Wortes *Brauch* zu einer Aktivierung der verwandten Information *Kultur*, *Tradition*, *Tracht* oder *Feier* führen. Auch wenn die Versuchsperson während des Vortests wahrscheinlich eine falsche Antwort rät (z. B. *Kultur*), sollte die korrekte Antwort (z. B. *Feier*) als Teil des Search-Sets mit aktiviert worden sein. Wird nun im Anschluss durch korrektives Feedback das korrekte Target-Wort bereitgestellt, sollte dieses Target-Wort durch die Voraktivierung besser enkodiert und die Verknüpfung zwischen dem Cue-Wort und dem Target-Wort dadurch gestärkt werden. Das sollte in einem finalen Test zu einer besseren Erinnerungsleistung für Target-Wörter der Vortest-Bedingung führen. Basierend auf den Annahmen der Search-Set-Hypothese sollte sich das Produzieren von Fehlern während eines Vortests nur dann positiv auf die spätere Erinnerungsleistung auswirken, wenn zwischen dem Lernmaterial (z. B. einem Cue-Wort und einem Target-Wort) eine semantische Assoziation besteht. Nur in diesem Fall ist davon auszugehen, dass das aktivierte Search-Set auch tatsächlich die korrekte Lösung beinhaltet. Würde beispielsweise unverwandte Cue-Target-Paare (z. B. *Hose* - *Rollladen*) als Lernmaterial verwendet werden, sollte die Präsentation des Cue-Wortes *Hose* zwar zu einer Aktivierung eines Search-Sets bestehend aus semantisch verwandter Information wie *Pullover*, *Klamotten* oder *Jeans* führen, aber nicht zu einer Voraktivierung der korrekten Lösung *Rollladen*. Ohne die Voraktivierung der korrekten Lösung gibt es daher keine Begründung, warum das korrekte Target-Wort bei der späteren Präsentation in der Vortest-Bedingung besser enkodiert und damit später besser abrufbar sein sollte als in der Lern-Bedingung, weshalb sich für dieses Material gemäß der Search-Set-Hypothese kein Vortest-Effekt zeigen sollte.

Insgesamt scheint es folglich eine Vielzahl an unterschiedlichen theoretischen Erklärungsansätzen für den Vortest-Effekt zu geben, von denen in diesem Abschnitt nur eine Auswahl prominenter Beispiele beschrieben wurde. Auch wenn diese Erklärungsansätze getrennt voneinander berichtet wurden, sind diese nicht zwingend als exklusive Erklärungsansätze zu verstehen und es wäre denkbar, dass ähnliche kognitive Mechanismen in mehreren der Theorien zum Tragen kommen (James, 2022). So gehen beispielsweise mehrere Erklärungsansätze davon aus, dass es während des Vortests, bzw. während des Vortest-Trials zu einer Aktivierung zusätzlicher Inhalte, wie beispielsweise semantisch verwandter Informationen (Elaborations-Hypothese) oder zeitlicher Kontextmerkmale

kommt (Kontext-Hypothese). Diese zusätzlichen Inhalte werden dann in das Cue-Target-Paar integriert und erleichtern in einem späteren finalen Test bei der Präsentation des Cue-Items den Abruf des Target-Items. Es sind also zusätzliche Verknüpfungen, ausgehend von dem Cue-Item, die gemäß diesen Ansätzen zu dem Erinnerungsvorteil durch Vortests führen.

Cue-Overload-Prinzip

Dies steht allerdings im Widerspruch zu den Annahmen des Cue-Overload-Prinzips (Raaijmakers & Shiffrin, 1981; Watkins & Watkins, 1975). Dieses Prinzip wurde ursprünglich zur Erklärung des Phänomens der Interferenz entwickelt, das den Befund beschreibt, dass die Erinnerung an Lerninhalte durch einen früheren Lernprozess gehemmt wird, was als proaktive Interferenz bezeichnet wird, oder durch das nachfolgende Lernen neuer Inhalte erschwert wird, was als retroaktive Interferenz bezeichnet wird. Diese Beeinträchtigung des Abrufs von Gedächtnisinhalten kommt gemäß dem Cue-Overload-Prinzip dadurch zustande, dass durch die zusätzlichen Lernprozesse mehrere Informationen mit einem Cue verknüpft werden, die in einem Abrufprozess miteinander konkurrieren. Dadurch reduziert sich die Wahrscheinlichkeit, dass die gesuchte Zielinformation abgerufen werden kann. Bekommen Versuchspersonen beispielsweise das Cue-Wort *Beruf* und das Target-Wort *Lehrer* präsentiert, entsteht eine Verknüpfung zwischen diesem Cue-Wort und dem Target-Wort und es ergibt sich so (abgesehen von vorher bestehenden Verknüpfungen mit dem Cue-Wort *Beruf*) in einem späteren Test nur die Abrufroute von dem Cue-Wort *Beruf* zu dem gesuchten Target-Wort *Lehrer*. Werden stattdessen entweder vor oder nach der Präsentation des Paares *Beruf* - *Lehrer* aber noch die Cue-Target-Paarungen *Beruf* - *Arzt* und *Beruf* - *Tischler* gelernt, kommt es zu einer Verknüpfung des Cue-Wortes *Beruf* mit mehreren Target-Wörtern, die in einem späteren Test miteinander konkurrieren und so den Abrufprozess erschweren, da das Cue-Wort überladen ist. In Bezug auf den Vortest-Effekt nimmt das Cue-Overload-Prinzip an, dass das Generieren eines Rateversuchs während eines Vortests zu einer Verknüpfung des Rateversuchs mit dem Cue-Item führt. Diese zusätzliche Verknüpfung führt zu einer Überladung des Cue-Items und bedingt in einem späteren finalen Test bei Präsentation des Cue-Items einen Antwortkonflikt, wenn die Versuchsperson versucht, die korrekte Lösung anstatt des eigenen Rateversuchs abzurufen. Gemäß dieser theoretischen Sichtweise sollten zusätzliche Verknüpfungen mit dem Cue-Item folglich den Abruf des Target-Items nicht erleichtern, sondern erschweren.

3. Offene Fragen und Ziele der vorliegenden Arbeit

Die bisherigen Studien zum Vortest-Effekt konnten bereits die Robustheit dieses Effekts für unterschiedliches Lernmaterial, verschiedene Testformate und auch in unterschiedlichen Untersuchungskontexten zeigen (für Überblick siehe z. B. Mera et al., 2022; Pan & Carpenter, 2023). Allerdings sind bisher noch einige Fragen, insbesondere bezüglich der optimalen Anwendung dieses Paradigmas, ungeklärt.

Während für andere Arten der Abrufübung bereits untersucht wurde, inwiefern die Anzahl an durchgeführten Tests einen Einfluss auf die Größe des daraus resultierenden Erinnerungsvorteils hat (z. B. Karpicke & Roediger, 2006), ist noch unklar, wie sich dieser Faktor auf die Größe des Vortest-Effekts auswirkt. So wäre es denkbar, dass wiederholte Rateversuche während eines Vortests zu einem noch größeren Erinnerungsvorteil führen oder aber diesen positiven Effekt reduzieren.

Des Weiteren finden sich in der Literatur zum Vortest-Effekt ausschließlich Studien, in denen das Lernen intentional stattfand, d. h. es erfolgte eine konkrete Lerninstruktion und die Versuchspersonen wurden darüber informiert, dass die präsentierten Inhalte zu einem späteren Zeitpunkt abgerufen werden sollen (z. B. Kornell et al., 2009). Noch offen ist hingegen die Frage, ob für die Entstehung des Vortest-Effekts ein gezielter Lernprozess erforderlich ist, oder ob sich dieser Erinnerungsvorteil auch dann zeigt, wenn das Lernen unbewusst und ohne direkte Lerninstruktion, also inzidentell, erfolgt.

Außerdem wurde in den meisten bisher durchgeführten Studien zum Vortest-Effekt im finalen Test das identische Material bzw. die identische Frage verwendet wie bereits im initialen Vortest. Im pädagogischen Alltag besteht ein finaler Test allerdings häufig auch aus Transfer-Fragen oder unbekannten Items, anhand derer der Lerninhalt abgerufen werden soll. Vor diesem Hintergrund ist die Frage, ob Lernende auch dann noch von Vortests profitieren, wenn in einem finalen Test neue, jedoch inhaltlich mit dem Lerninhalt verwandte Fragen oder Cue-Items eingesetzt werden, insbesondere für die effektive Anwendung des Vortest-Paradigmas im Bildungskontext von großer Relevanz.

Lernprozesse finden vor allem in der Praxis häufig unter Ablenkung statt, beispielsweise durch Mitlernende oder Hintergrundgeräusche, wodurch die Aufmerksamkeit für die Lerninhalte beeinträchtigt sein kann (z. B. Unsworth & McMillan, 2017). Zur

kontrollierten Untersuchung des Einflusses von Ablenkung und reduzierter Aufmerksamkeit auf die Lernleistung wird in der experimentellen Forschung typischerweise das Paradigma der geteilten Aufmerksamkeit verwendet. Dabei findet parallel zu einer Lernaufgabe noch eine weitere Aufgabe statt, die zu einer Konkurrenz um Aufmerksamkeitsressourcen führt (für Überblick siehe Mulligan, 2008). Während andere Arten der Abrufübungen bereits unter geteilter Aufmerksamkeit untersucht wurden (z. B. Mulligan und Picklesimer, 2016), ist dieser Einfluss für das Vortest-Paradigma bislang noch unerforscht. Für den gezielten Einsatz von Vortests im Lernalltag ist es jedoch von großem Interesse, ob und in welchem Ausmaß Aufmerksamkeitsressourcen während des initialen Vortests oder während der Phase des korrektiven Feedbacks für die Entstehung des Vortest-Effekts erforderlich sind.

Insgesamt bestehen für das Paradigma des Vortest-Effekts also noch zahlreiche empirische Fragen, insbesondere hinsichtlich möglicher Einflussfaktoren für eine effiziente Anwendung von Vortests. Das Ziel der vorliegenden Arbeit ist es daher, anhand der neun durchgeführten Experimente einen wichtigen Beitrag zur Schließung der vorhandenen Lücken zu leisten. So untersuchen die Experimente 1-4 den Einfluss wiederholter Rateversuche während des Vortests, um die Frage nach der gewinnbringendsten Anzahl an Vortests zu beantworten. Nachfolgend widmen sich die Experimente 5-7 der Fragestellung, ob auch inzidentelles Lernen einen Vortest-Effekt hervorrufen kann. Zudem soll mit diesen Experimenten geklärt werden, ob im finalen Test die Präsentation des ursprünglichen Cue-Items notwendig ist, um zu einer verbesserten Erinnerungsleistung zu führen, oder ob sich auch unter Verwendung semantisch verwandter Mediator-Cues ein Vortest-Effekt finden lässt. Die abschließenden Experimente 8 und 9 betrachten den Vortest-Effekt unter geteilter Aufmerksamkeit, um zu prüfen, ob und in welcher der Phasen des Vortest-Trials eine Ablenkung durch eine Zweitaufgabe die Erinnerungsleistung und somit die Größe des Vortest-Effekts beeinflusst.

Neben den beschriebenen empirischen Lücken bleiben in der bisherigen Literatur vor allem die theoretischen Hintergründe und insbesondere die Frage, welche kognitiven Mechanismen zur Entstehung des Vortest-Effekts beitragen, noch ungeklärt. Zwar werden unterschiedliche Erklärungsansätze für den Vortest-Effekt diskutiert, bisher wurden aber nur selten konkrete Annahmen einzelner Erklärungsansätze experimentell untersucht. Ein weiteres Ziel der vorliegenden Arbeit war es daher, erste Hinweise auf die Bedeutung

unterschiedlicher kognitiver Prozesse für den Vortest-Effekt zu gewinnen, indem die durchgeführten Experimente auch die Vorhersagen verschiedener Theorien überprüfen.

Die Ergebnisse der Experimente 1-4 sollten Erkenntnisse darüber liefern, ob das wiederholte Generieren eines Rateversuchs zu einem zusätzlichen Vorteil führt oder ob mit zunehmender Anzahl an Rateversuchen der Erinnerungsvorteil von Vortests verschwindet. Letzteres würde dafürsprechen, dass es durch die Verknüpfung der zusätzlichen Rateversuche mit dem Cue-Item zu einer Überladung dieses Items kommt und dadurch der Abruf der korrekten Lösung im finalen Test beeinträchtigt ist, womit diese Experimente konkrete Annahmen des Cue-Overload-Prinzips testen. Sollte sich in den Experimenten 1-4 hingegen ein größerer Vortest-Effekt nach wiederholten Rateversuchen als nach nur einem Rateversuch zeigen, würde dies die Annahme stützen, dass Elaborations- und Aufmerksamkeitsprozesse eine wichtige Rolle für die Entstehung dieses Effektes spielen.

Anhand der nachfolgenden Experimente 5-7 können dann vor allem Aussagen über die Error-Correction-Hypothese und die Elaborations-Hypothese getroffen werden. So würde ein ausbleibender Vortest-Effekt unter Verwendung semantischer Mediatoren als Cues dafür sprechen, dass es durch Fehlerkorrekturmechanismen zu einer Stärkung der zielführenden Abrufroute, aber zugleich auch zu einer Unterdrückung anderer Verknüpfungen des Lernmaterials kommt. Würde sich hingegen auch für diese Cue-Items ein Vortest-Effekt zeigen, würde dies die konkrete Annahme der Elaborations-Hypothese, wonach Vortests zu einer Aktivierung eines semantischen Netzwerkes führen und das Lernmaterial in dieses eingebettet wird, stützen.

Die abschließenden Experimente 8 und 9, in denen der Vortest-Effekt unter geteilter Aufmerksamkeit untersucht wird, erlauben vor allem die gezielte Überprüfung der Aufmerksamkeits-Hypothese. Denn dieser Erklärungsansatz würde annehmen, dass insbesondere während der Phase des korrektiven Feedbacks die Verfügbarkeit von Aufmerksamkeitsressourcen für die Entstehung des Vortest-Effekts eine zentrale Rolle spielt. Sollte sich also in den Experimenten 8 und 9 eine geteilte Aufmerksamkeit durch Ablenkung vor allem in dieser Phase negativ auf den Vortest-Effekt auswirken, würde dies die Rolle von Aufmerksamkeitsprozessen für diesen Effekt bekräftigen. Insgesamt sollen mit der vorliegenden Arbeit also einerseits empirische Fragen beantwortet und andererseits theoretische Erkenntnisse über den Vortest-Effekt gewonnen werden.

4. Der Einfluss wiederholter Tests auf die Erinnerungsleistung

4.1. Bisherige Befunde

Wie die vorherigen Abschnitte zeigten, führt die Bearbeitung von Tests sowohl nach dem Lernen, zwischen dem Lernen mehrerer Listen und auch bereits vor dem Lernen der zu erinnernden Inhalte zu einem späteren Erinnerungsvorteil für die relevante Information. Daraus leitet sich die Frage ab, wie oft ein Test durchgeführt werden sollte, um diesen Vorteil zu maximieren. Hilft hier viel auch viel oder ist viel mehr die Tatsache entscheidend, dass Lernende grundsätzlich eine Abrufübung durchführen?

Diese Fragestellung untersuchten beispielsweise Roediger und Karpicke (2006) in ihrer Studie, wobei sie das klassischen Test-Effekt Paradigma verwendeten. Dabei sollten Studierende den Inhalt eines Prosa-Textes lernen und hatten dabei die Möglichkeit, diesen Text viermal in Folge erneut zu lernen, diesen Text dreimal erneut zu lernen und abschließend eine einzige Abrufübung durchzuführen oder den Text nach einmaligem Lernen im Anschluss dreimal abzurufen, wobei die Abrufübung in Form eines freien Abrufs konzipiert wurde. In einem späteren finalen Test zeigte sich einerseits eine signifikant bessere Erinnerungsleistung in der Bedingung mit nur einem Test verglichen mit der erneuten Lern-Bedingung, was den klassischen Test-Effekt darstellt. Vor allem aber ergab sich ein besserer Abruf, wenn drei Abrufübungen stattfanden verglichen mit der einmaligen Abruf-Bedingung. Dies legt nahe, dass Lernende von wiederholten Tests noch stärker zu profitieren scheinen (Roediger & Karpicke, 2006). Diese Zunahme der Größe des klassischen Test-Effekts von der Durchführung eines einzelnen Tests auf die wiederholte Durchführung von Tests konnte auch in weiteren Studien repliziert werden (z. B. Karpicke & Roediger, 2008).

Eine weitere Möglichkeit, den Effekt wiederholter Abrufübungen auf die spätere Gedächtnisleistung zu untersuchen, stellt die Verwendung eines Kriteriumlevels dar, wobei die Anzahl korrekter Abrufdurchgänge während der Test-Phase festgelegt wird. So sollten

Versuchspersonen in der Studie von Vaughn und Rawson (2011) beispielsweise Englisch-Litauisch Vokabeln lernen, die in einer anschließenden Übungs-Phase in einem Cued-Recall-Test abgerufen werden sollten. Dabei sollten die Target-Wörter entweder einmal, zweimal, dreimal, viermal oder fünfmal korrekt abgerufen werden, bevor die Übungs-Phase für das entsprechende Target-Wort beendet war. Folglich durchlief ein Teil der Target-Wörter nur eine Abrufübung, wohingegen für andere Target-Wörter bis zu fünf Abrufübungen durchgeführt wurden. Die Ergebnisse zeigten dabei eine Zunahme der Erinnerungsleistung mit steigender Anzahl an Tests. Ähnliche Befunde liefern auch die Studien von Karpicke und Roediger (2007) und Pyc und Rawson (2009), in deren Experimenten ebenfalls eine bessere Erinnerungsleistung nach wiederholten Tests verglichen mit nur einem Test gefunden werden konnte, wobei die Lerninhalte bis zu achtmal (Karpicke & Roediger, 2007) bzw. bis zu zehnmal (Pyc und Rawson, 2009) korrekt abgerufen werden sollten. Eine Meta-Analyse von Yang et al. (2021) betrachtete diesen Einfluss der Test-Häufigkeit auf die spätere Erinnerungsleistung über mehr als 500 Studien hinweg, wobei die Autoren eine Unterteilung in vier Kategorien (ein Test, zwei Tests, drei oder mehr Tests, unbegrenzte Anzahl an Tests) vornahmen. Dabei zeigte sich eine signifikante Zunahme der Erinnerungsleistung mit steigender Anzahl an Tests, die sich auch mit einer Regressionsanalyse bestätigen ließ, bei der sich eine Zunahme der Erinnerungsleistung mit jedem weiteren Test ergab. Lernende scheinen also einen zusätzlichen Vorteil aus der wiederholten Durchführung von Tests zu erhalten und dieser Effekt scheint unabhängig von der Schwierigkeit des verwendeten Materials zu sein (deJonge & Tabbers, 2013).

Eine mögliche Erklärung dafür liefern McDaniel et al. (1989), die argumentierten, dass die Bearbeitung von Tests das Lernen und damit die spätere Erinnerungsleistung für die gelernten Inhalte fördert, indem diese zu einer Verstärkung bestehender Gedächtnisspuren sowie einer Förderung der Cue-Target-Beziehung führen. Dieser Gedanke wurde später von Carpenter (2009) in der Elaborations-Hypothese aufgegriffen, nach der die Verknüpfung zwischen dem Cue-Wort und dem Target-Wort dadurch gefördert wird, dass es durch die Abrufübung zu einer Aktivierung semantisch verwandter Information kommt, die die Cue-Target-Verknüpfung inhaltlich zusätzlich anreichert und so zu einer detaillierten Gedächtnisspur führt. Gemäß dieser Sichtweise könnte jeder zusätzliche Test neue, unterschiedliche Informationen aktivieren, die zu weiteren Verknüpfungen führen und so zu

einem späteren Zeitpunkt den Abruf der Target-Information erleichtern. Zusätzlich könnten Tests auch dazu führen, dass sich die Anzahl der möglichen Abrufrouen zur Zielinformation erhöht, weshalb auch hier wiederholte Tests zu einer noch besseren Erinnerungsleistung beitragen könnten als nur ein einziger Test (Bjork, 1975, 1988). In der Literatur zum klassischen Test-Effekt finden sich folglich vielfältige Studien, die den Einfluss der Testhäufigkeit auf den späteren Erinnerungsvorteil untersuchten. Für die Verwendung des Vortest-Paradigmas lassen sich bisher jedoch nur vereinzelte Befunde zur Untersuchung des Einflusses wiederholter Tests auf die Größe des Vortest-Effekts finden.

Eine Studie stammt dabei von Vaughn und Rawson (2012), in deren Experiment 1 Versuchspersonen schwach assoziierte Cue-Target-Paare lernen sollten, die entweder direkt vollständig präsentiert und gelernt wurden (Lern-Bedingung) oder für die zunächst nur das Cue-Wort zu sehen war und das korrekte Target-Wort erraten werden sollte (Vortest-Bedingung). Dabei sollten die Versuchspersonen pro Wortpaar immer entweder einen einzigen Rateversuch abgeben oder pro Cue-Wort drei Rateversuche generieren. Erst nachdem für alle Items der Vortest bearbeitet wurde, folgte das korrektive Feedback, wobei alle Wortpaare vollständig präsentiert wurden und gelernt werden konnten. Abschließend sollte für alle Items ein finaler Cued-Recall-Test bearbeitet werden, bei dem sich eine bessere Erinnerungsleistung für Target-Wörter der Vortest-Bedingung mit drei Rateversuchen verglichen mit der Vortest-Bedingung mit nur einem Rateversuch zeigte. Für keine der beiden Rate-Bedingungen konnten die Autoren allerdings einen Vortest-Effekt nachweisen. Im Gegenteil: Target-Wörter, die in der Lern-Bedingung präsentiert wurden, konnten besser abgerufen werden (Vaughn & Rawson, 2012). Jedoch wurde in der beschriebenen Studie das korrektive Feedback und somit die Lern-Phase in der Vortest-Bedingung nicht unmittelbar nach dem Vortest, sondern erst mit einer zeitlichen Verzögerung dargeboten. Diese Prozedur führte allerdings wiederholt zu einem Ausbleiben des Vortest-Effekts (z. B. Grimaldi & Karpicke, 2012; Hays et al., 2013), weshalb es nicht überraschend erscheint, dass in der beschriebenen Studie ebenfalls kein Vortest-Effekt gefunden werden konnte. Aufgrund des fehlenden Vortest-Effekts gilt es auch den Befund der besseren Erinnerungsleistung nach wiederholten Vortests verglichen mit nur einem Vortest mit Vorsicht zu interpretieren.

James (2022) untersuchte ebenfalls den Einfluss wiederholter Tests auf die Größe des Vortest-Effekts. In drei Experimenten bekamen Versuchspersonen Wissensfragen

präsentiert (z. B. „Welche Milch eines Tieres gerinnt nicht?“), die entweder direkt mit der korrekten Antwort (z. B. „Kamel“) zu sehen waren (Lern-Bedingung) oder für die zunächst ein einziger oder drei Rateversuche abgegeben werden sollten (Vortest-Bedingungen), bevor unmittelbar danach die korrekte Lösung präsentiert wurde. In einem abschließenden Cued-Recall-Test wurden den Versuchspersonen alle Fragen erneut dargeboten und es sollte die dazugehörige korrekte Antwort gegeben werden. In allen drei Experimenten zeigte sich bei Betrachtung der Ergebnisse ein signifikanter Vortest-Effekt zwischen der Lern-Bedingung und der Vortest-Bedingung mit nur einem Rateversuch. Zwischen der Erinnerungsleistung der beiden Vortest-Bedingungen hingegen ließ sich kein Unterschied finden, d. h. sowohl ein einziger Rateversuch als auch drei Rateversuche produzierten in dieser Studie einen vergleichbaren Erinnerungsvorteil gegenüber der Lern-Bedingung. Allerdings stellt sich hier die Frage, ob sich dieser Befund auch für unterschiedliches Lernmaterial finden lässt.

4.2. Theoretische Erwartungen

Die bisherige Befundlage zur Untersuchung der Testanzahl auf den Vortest-Effekt ist gemischt und lässt sich dementsprechend nur bedingt interpretieren. Allerdings könnte die Betrachtung dieses Aspekts Rückschlüsse über mögliche zugrundeliegende Mechanismen dieses Gedächtnis-Effekts erlauben, da hier verschiedene theoretische Erklärungsansätze unterschiedliche Ergebnisse vermuten lassen würden. Deshalb sollen anschließend verschiedene Annahmen näher beschrieben werden.

Das Cue-Overload-Prinzip (Raaijmakers & Shiffrin, 1981; Watkins & Watkins, 1975) nimmt beispielsweise an, dass es während des Vortests durch die Präsentation eines Cue-Items (Haus) und das Generieren eines Rateversuchs (*Türe*) zu einer Verknüpfung des Cue-Items mit dem Rateversuch (Haus – *Türe*) kommt. Wird nachfolgend das Cue-Item zusammen mit dem korrekten Target-Item (Haus - *Fenster*) präsentiert, entsteht auch hier eine Verknüpfung ausgehend von dem Cue-Item, wodurch das Cue-Item nicht mehr nur mit einem Item, sondern mit zwei Items assoziiert ist, d. h. das Cue-Item ist überladen. Wird in einem späteren finalen Cued-Recall-Test das Cue-Item (Haus) präsentiert, führt dies zu einem Antwortkonflikt zwischen der korrekten Abrufroute (Haus – *Fenster*) und der ebenfalls während des Vortests entstandenen Abrufroute zum eigenen Rateversuch (Haus – *Türe*), wodurch der Abruf des Target-Items erschwert wird. Gemäß diesem Erklärungsansatz

wäre es denkbar, dass wiederholte Vortests zu einer Verschärfung dieses Antwortkonflikts beitragen. So sollte jeder zusätzliche Vortest und der damit einhergehende Rateversuch mit dem Cue-Item verknüpft werden und dadurch zu einem zusätzlichen Konkurrenten werden, der im finalen Test möglicherweise anstelle der korrekten Target-Information abgerufen wird. Daher würde das Cue-Overload-Prinzip zu der Erwartung führen, dass wiederholte Vortests mit einer schlechteren Erinnerungsleistung im finalen Test verglichen mit nur einem einzigen Vortest einhergehen würden, da diese zu einer stärkeren Überladung des Cue-Items führen.

Eine gegensätzliche Erwartung ergibt sich hingegen aus der Elaborations-Hypothese des Vortest-Effekts (Carpenter, 2009, 2011), die davon ausgeht, dass das Produzieren falscher Antworten im Vortest den späteren Abruf der Zielinformation nicht erschwert, sondern diesen sogar fördert. Wie zuvor beschrieben, nimmt dieser Erklärungsansatz an, dass es während der Bearbeitung eines Vortests durch die Präsentation eines Cue-Items zu einer Aktivierung eines semantischen Netzes aus verwandter Information kommt, die mit dem Cue-Target-Paar verknüpft wird, sobald das korrektive Feedback bereitgestellt wird. Diese zusätzlichen Informationen, einschließlich des falschen Rateversuchs, sollen dann in einem späteren finalen Test durch die erneute Darbietung des Cue-Items reaktiviert werden und im Sinne eines semantischen Mediators den Abruf des Target-Items erleichtern (Huelser & Metcalfe, 2012). Basierend auf der Elaborations-Hypothese wäre folglich zu erwarten, dass die wiederholte Durchführung von Vortests die Erinnerungsleistung für das Lernmaterial zusätzlich fördert. Denn werden beispielsweise die Versuchspersonen dazu aufgefordert, statt nur eines einzigen Rateversuchs, zwei oder drei Rateversuche zu generieren, sollte dies zu einer noch vielfältigeren und stärkeren Aktivierung von Informationen, die semantisch mit dem Cue-Item verwandt sind, führen. Außerdem könnten die fehlerhaften Rateversuche als zusätzliche Mediatoren fungieren, die im finalen Test den Abruf des Target-Items erleichtern, wodurch wiederholte Vortests im Vergleich zu nur einem einzigen Vortest zu einem noch größeren Vortest-Effekt führen könnten.

Auch die Aufmerksamkeits-Hypothese des Vortest-Effekts würde zu einer ähnlichen Annahme führen. Dieser Erklärungsansatz geht davon aus, dass es durch das Bearbeiten eines Vortests zu der Feststellung einer Diskrepanz zwischen aktuellem und erforderlichem Wissenstand kommt. Dadurch soll eine gesteigerte Neugierde und Motivation entstehen, das

Target-Item zu lernen, weshalb mehr Aufmerksamkeit auf die Verarbeitung des Target-Items gelenkt wird (Potts et al., 2019). Wenn die Versuchspersonen nun mehrere Rateversuche während des Vortests abgeben, könnte dies zu einer noch gesteigerten Neugierde und Aufmerksamkeit und somit zu einer noch besseren Erinnerungsleistung für diese Target-Items führen, verglichen mit nur einem Rateversuch.

Die Kontext-Hypothese des Vortest-Effekts geht davon aus, dass es während des Enkodierens zu einer Verknüpfung der Lerninhalte mit den dabei präsenten räumlich-zeitlichen Kontextmerkmalen kommt (z. B. Mensik & Raaijmakers, 1988). Diese Kontextmerkmale werden in einem finalen Test reaktiviert und sollen somit den Zugang zur Zielinformation erleichtern. Während Items in der Lern-Bedingung nur mit dem Lern-Kontext verknüpft sind, sollten Items in der Vortest-Bedingung ebenfalls mit dem Lern-Kontext, zusätzlich aber auch noch mit den Merkmalen des Vortest-Kontexts assoziiert sein. Items in einer Vortest-Bedingung mit wiederholten Rateversuchen hingegen könnten mit einem noch umfangreicheren und vielfältigeren Kontext verbunden sein, da hier nicht nur die Merkmale aus einem Vortest-Kontext, sondern aus mehreren unterschiedlichen Vortest-Kontexten integriert sein sollten. Diese größere Vielfalt hilfreicher Kontextmerkmale könnte die Wahrscheinlichkeit erhöhen, dass es durch die Reaktivierung eines oder mehrerer Kontextmerkmale zu einem korrekten Abruf der Zielinformation kommt, und somit zu einer noch besseren Erinnerungsleistung beitragen. Die Größe des Vortest-Effekts könnte folglich gemäß der Kontext-Hypothese mit steigender Anzahl an Rateversuchen noch zunehmen.

Die Search-Set-Hypothese (Grimaldi & Karpicke, 2012) nimmt an, dass es während des Vortests durch die Präsentation des Cue-Items zu einer Aktivierung eines Search-Sets kommt, das die korrekte Antwort beinhaltet und, dass es durch diese Voraktivierung zu einer tieferen Enkodierung und später einem besseren Abruf des Target-Items kommt. Sollen statt nur eines Rateversuchs mehrere Rateversuche während des Vortests abgegeben werden, könnte dies zur Aktivierung mehrerer, leicht unterschiedlicher Search-Sets führen. Dadurch sollte sich die Wahrscheinlichkeit, dass das korrekte Target-Item innerhalb eines der Search-Sets voraktiviert wird, erhöhen, weshalb wiederholte Rateversuche zu einer noch besseren Enkodierung und einem leichteren Abruf der Zielinformation führen könnten. Auch gemäß diesem Erklärungsansatz wäre also ein noch größerer Vortest-Effekt bei wiederholten Rateversuchen verglichen mit nur einem Rateversuch zu erwarten.

Auf Grundlage der Error-Correction-Hypothese (Carrier & Pashler, 1992) sollte es nach der Abgabe eines fehlerhaften Rateversuchs bei der nachfolgenden Präsentation der korrekten Lösung zur Entstehung eines Fehlersignals kommen. Dieses Fehlersignal sollte erst nach dem Vortest in der Phase des korrektiven Feedbacks ausgelöst werden, weshalb sich die Anzahl an Rateversuchen während des Vortests nicht auf die Entstehung dieses Signals auswirken sollte. Da es aber vor allem die durch das Fehlersignal ausgelösten Fehlerkorrekturmechanismen sind, die gemäß diesem Erklärungsansatz zur Entstehung des Vortest-Effekts führen, sollte die Größe dieses Effekts nicht in Abhängigkeit der Anzahl der Rateversuche variieren.

Insgesamt wäre es also denkbar, dass wiederholte Vortests zu einer Beeinträchtigung der Erinnerungsleistung und zu einem Ausbleiben des Vortest-Effekts führen, sich die Anzahl an Vortests nicht auf die Größe des Effekts auswirkt, oder aber wiederholte Rateversuche zu einem zusätzlichen Erinnerungsvorteil und damit zu einem Anstieg des Vortest-Effekts beitragen könnten.

4.3. Fragestellung der Experimente 1-4

Die bisherige Literatur zur Untersuchung des Einflusses wiederholter Tests auf die spätere Erinnerungsleistung betrachtet nahezu ausschließlich das Paradigma des klassischen Test-Effekts, wobei Versuchspersonen von wiederholten Tests zu profitieren scheinen. Allerdings gibt es noch kaum Studien, die den Effekt der Anzahl an Vortests, in denen die Erfolgsquote deutlich niedriger ist als in den interpolierten Tests im Test-Effekt-Paradigma, auf die spätere Erinnerung an die relevante Zielinformation untersuchten.

Bisherige Studien über die Verwendung von Vortests als Lernstrategie konnten bereits einen vorteilhaften Effekt von Vortests auf die Erinnerungsleistung sowohl im Labor-Setting (z. B. St. Hilaire & Carpenter, 2020; Kliegl et al., 2024), als auch im angewandten Klassenzimmer-Setting (Carpenter et al., 2018) für einen Zeitraum von bis zu einer Woche zeigen. Jedoch werfen diese Befunde die Frage auf, ob Lehrpersonen bei der Verwendung von Vortests möglicherweise noch mehr Zeit investieren sollten. Beispielsweise könnte eine Lehrperson, bevor ein spezifisches Thema im Unterricht behandelt wird (z. B. Die Wirtschaft eines Landes), die Lernenden dazu auffordern, nicht nur eine einzige Antwort auf eine Vortest-Frage zu diesem Thema zu geben (z. B. „Welche Produkte exportiert Dänemark?“),

sondern mehrfach darauf zu antworten. Ob sich ebendieser zusätzliche Aufwand bei der Verwendung von Vortests auszahlt und zu einem noch größeren Vorteil führt oder nicht, ist daher von entscheidender Bedeutung für die praktische Anwendung. Die Untersuchung des Einflusses der Anzahl an Rateversuchen auf die Größe des Vortest-Effekts dient also einerseits dazu, diese bestehende empirischen Lücke zu schließen und Implikationen für die Anwendung des Paradigmas im pädagogischen Kontext abzuleiten.

Andererseits ermöglicht die Beantwortung der Frage, wie sich wiederholte Rateversuche auf die Größe des Vortest-Effekts auswirken, auch wichtige Erkenntnisse über die theoretischen Hintergründe dieses Gedächtniseffekts. So sollten die Ergebnisse dieser Experimentserie Hinweise darüber liefern, ob das wiederholte Generieren eines Rateversuchs zu zusätzlichen Verknüpfungen mit dem Cue-Item führt und ob sich diese förderlich oder hinderlich auf den Vortest-Effekt auswirken. Daraus würden sich konkrete Aussagen über die zugrunde liegenden Mechanismen ableiten lassen.

Im nachfolgenden Kapitel werden die Ergebnisse von vier Experimenten berichtet, welche durchgeführt wurden, um zu untersuchen, ob wiederholte Vortests zu einer Vergrößerung des Vortest-Effekts führen können und falls ja, ob sich dieser Einfluss auch auf die Verwendung unterschiedlich komplexer Materialien generalisieren lässt. Dazu wurde in den Experimenten 1-3 eine typische Aufgabe zur Untersuchung des Vortest-Effekts verwendet, bei der die Versuchspersonen Cue-Target-Paare lernen sollten. Diese Cue-Target-Paare wurden entweder über die gesamte Dauer des Lern-Trials vollständig präsentiert und konnten dementsprechend von Beginn des Lern-Trials an gelernt werden oder es wurde zunächst nur das Cue-Wort präsentiert und das dazugehörige Target-Wort sollte erraten werden, bevor das vollständige Paar zu sehen war (z. B. Kornell et al., 2009). Anders als im klassischen Vortest-Paradigma sollten die Versuchspersonen dabei jedoch im initialen Vortest nach der Abgabe ihres Rateversuchs noch einen (Experiment 1) oder zwei weitere (Experimente 1-3) Rateversuche abgeben. In einem späteren finalen Test wurde die Gedächtnisleistung an alle Wortpaare mittels eines Cued-Recall-Tests bestimmt, bei dem immer das Cue-Wort eines Wortpaares präsentiert und das korrekte, dazugehörige Target-Wort abgerufen werden sollte. Dabei wurden in den Experimenten 1 und 3 schwach assoziierte Wortpaare und in Experiment 2 Swahili-Deutsch-Vokabeln als Lernmaterial verwendet. In den Experimenten 1 und 2 nahm die Länge der Lern-Trials mit der Anzahl an

Rateversuchen zu, während die Dauer aller Vortest-Trials in Experiment 3 konstant gehalten wurde. Die konstante Dauer in Experiment 3 wurde gewählt, um auszuschließen, dass die Zeit, die während des Ratens zur Verfügung steht, für mögliche Veränderungen in der Größe des Vortest-Effekts verantwortlich ist. In Experiment 4 sollten die Versuchspersonen drei Textpassagen zu unterschiedlichen Inhalten (z. B. über das Land Brasilien) lernen, wobei immer eine der Textpassagen von Beginn an vollständig zu lesen war. Die beiden anderen Textpassagen enthielten zunächst noch Lücken in den Sätzen, für die die Versuchspersonen entweder mit einem Versuch oder mit drei Versuchen die fehlende Lösung erraten sollten, bevor die korrekte Information präsentiert wurde.

Für alle vier Experimente wurde ein klassischer Vortest-Effekt beim Vergleich der Vortest-Bedingung mit einem Rateversuch und der Lern-Bedingung erwartet (z. B. Kornell et al., 2009; Richland et al., 2009). Unklar war hingegen, ob das wiederholte Bearbeiten eines Vortests zu einem Anstieg oder einer Reduktion des Vortest-Effekts führt. Während sowohl die Elaborations-Hypothese und die Kontext-Hypothese als auch das Cue-Overload-Prinzip annehmen, dass wiederholte Vortests zu einer verstärkten Gedächtnisspur verknüpft mit der Cue-Information führen, ergeben sich daraus jedoch unterschiedliche Vorhersagen bezüglich der Auswirkungen dieser zusätzlich verknüpften Informationen auf den Abruf der Target-Information. Gemäß dem Cue-Overload-Prinzip sollte diese zusätzliche Information eine Quelle für Interferenz darstellen, die den Abruf der Target-Information beeinträchtigt, wohingegen die Elaborations-Hypothese und die Kontext-Hypothese davon ausgehen, dass die zusätzliche Information den Zugang zur Target-Information erleichtern sollte. Auch die Aufmerksamkeits-Hypothese und die Search-Set-Hypothese würden erwarten, dass wiederholte Rateversuche zu einem verbesserten Abruf der Target-Items beitragen.

Ein drittes denkbare Szenario wäre, dass die Anzahl an Rateversuchen keinen Einfluss auf die Größe des Vortest-Effekts hat, wie dies möglicherweise auf Grundlage der Error-Correction-Hypothese anzunehmen wäre. Die nachfolgenden Experimente sollen dementsprechend erste Erkenntnisse darüber liefern, ob sich das wiederholte Generieren eines Rateversuchs im initialen Vortest positiv, negativ oder gar nicht auf die Größe des Vortest-Effekts auswirkt.

5. Der Vortest-Effekt mit wiederholten Rateversuchen beim Test

5.1. Experiment 1

Methode

Ethische Berücksichtigungen. Alle nachfolgend berichteten Experimente 1-4 wurden in Übereinstimmung mit den Vorgaben der Deklaration von Helsinki des Weltärztebundes durchgeführt.

Stichprobe. Um die Stichprobengröße für die Experimente 1-4 zu ermitteln, wurde eine Poweranalyse mit dem Programm G*Power (Faul et al., 2009) durchgeführt, wobei das α auf einen Wert von .05, das β auf einen Wert von .20 und die Effektgröße basierend auf der metaanalytischen Schätzung der Größe des Vortest-Effekts auf ein Hedges' g von 0.44 (Boustani & Shanks, 2022) festgesetzt wurde. Dabei ergab sich eine notwendige Stichprobengröße von $N = 43$. Dieser Empfehlung folgend nahmen am Experiment 48 Studierende ($M = 24.5$ Jahre, Altersspanne = 18-34 Jahre, 40 weiblich) aus unterschiedlichen Universitäten in Deutschland teil. Für alle Versuchspersonen war die Muttersprache Deutsch und alle Versuchspersonen gaben zu Beginn des Experimentes ihr mündliches Einverständnis. Für die Teilnahme wurden entweder Versuchspersonenstunden im Rahmen des Psychologiestudiums gutgeschrieben oder eine monetäre Entlohnung vergeben.

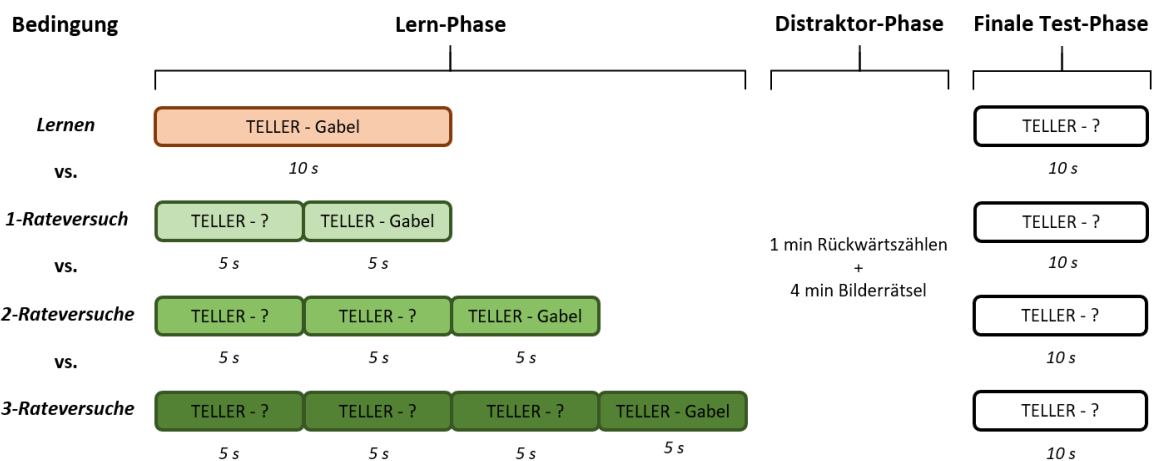
Material. Als Lernmaterial wurden 60 schwach miteinander verwandte Wortpaare (z. B. Teller - *Gabel*, Tradition - *Weihnachten*) verwendet, die aus den Normen von Nelson et al. (1998) entnommen und ins Deutsche übersetzt wurden. Die vorwärtsgerichtete Assoziationsstärke der Wortpaare lag dabei zwischen .051 und .053 und betrug im Mittel einen Wert von .052. Das bedeutet, dass in der Ursprungsstudie von Nelson et al. (1998) die Versuchspersonen in etwa 5 % der Fälle das korrekte Target-Wort als erste Antwort gaben, wenn ihnen ein Cue-Wort präsentiert wurde. Die Wortpaare wurden in vier Untergruppen

mit jeweils 15 Wortpaaren unterteilt, wobei jede Untergruppe gleich häufig in jeder der Experiment-Bedingungen verwendet wurde.

Design und Versuchsablauf. Das gesamte Experiment wurde online mittels der Videotelefonie-Software Zoom (Zoom Video Communications Inc., 2016) durchgeführt, wobei alle Testungen in einzelnen Videokonferenzen stattfanden. Das Experiment bestand dabei aus drei Phasen, die von jeder Versuchsperson durchlaufen wurden (siehe Abbildung 1): einer Lern-Phase, einer Distraktor-Phase und einer finalen Test-Phase.

Abbildung 1

Experimentelles Design des Experimentes 1



Anmerkung: Unterteilung des Versuchsablaufes in Lern-Phase, Distraktor-Phase und finale Test-Phase, wobei jede Versuchsperson in der Lern-Phase die identische Anzahl an Items in jeder der vier Bedingungen präsentiert bekam. Im finalen Test wurden alle gelernten Wörter mit einem Cued-Recall-Test abgefragt.

In der Lern-Phase wurden den Versuchspersonen alle 60 Wortpaare der Reihe nach in der Bildschirmmitte dargeboten, von denen allerdings nur die 15 Wortpaare der Lern-Bedingung von Beginn an vollständig präsentiert wurden. In den Trials dieser Lern-Bedingung wurde das Cue-Wort zusammen mit dem Target-Wort für 10 s präsentiert (z. B. Teller - Gabel) und konnte somit direkt gelernt werden. Für die anderen 45 Wortpaare sollten die Versuchspersonen zunächst einen Vortest bearbeiten. Dabei wurden jeweils 15 Wortpaare der 1-Rateversuch-Bedingung, der 2-Rateversuche-Bedingung sowie der 3-

Rateversuche-Bedingung zugeteilt. In der 1-Rateversuch-Bedingung wurde für 5 s zunächst nur das Cue-Wort präsentiert und die Versuchspersonen sollten versuchen, das dazugehörige Target-Wort zu erraten (z. B. Tradition - ?). Dabei war zu Beginn eines jeden Trials für 0.5 s ein Ton zu hören, der als Signal dafür fungierte, dass nun der Rateversuch abgegeben werden sollte. In der 2-Rateversuche-Bedingung und der 3-Rateversuche-Bedingung folgte daraufhin eine bzw. zwei weitere 5 s-Phasen, in denen die Teilnehmer einen bzw. zwei weitere Rateversuche abgeben sollten. Dabei wurde zu Beginn jeder dieser weiteren 5 s-Phasen erneut der Signalton für 0.5 s präsentiert und die Versuchspersonen wurden vorab instruiert nach jedem Ton einen neuen, unterschiedlichen Rateversuch abzugeben. Alle Rateversuche sollten dabei mündlich abgegeben werden. Die Versuchspersonen wussten zu Beginn eines jeden Vortest-Trials allerdings nicht, wie viele Rateversuche nachfolgend abgegeben werden sollten. Am Ende eines jeden Vortest-Trials wurde den Versuchspersonen unabhängig von der Anzahl der abzugebenden Rateversuche das vollständige Wortpaar (z. B. Tradition - *Weihnachten*) für weitere 5 s präsentiert. In der Summe betrug die Dauer eines Trials in der Lern- und der 1-Rateversuch-Bedingung somit 10 s, in der 2-Rateversuche-Bedingung 15 s und in der 3-Rateversuche-Bedingung 20 s. Alle Wortpaare wurden jeder Versuchsperson in einer zufälligen Reihenfolge präsentiert.

Nach der Lern-Phase folgte die Distraktor-Phase, in der die Versuchspersonen aufgefordert wurden, zunächst für eine Minute laut in 7er-Schritten rückwärtszuzählen, bevor für weitere 4 min in einem Online-Spiel die Unterschiede zwischen zwei Bildern gesucht werden sollten (<https://www.suchbilder.com/fehlerbilder/>).

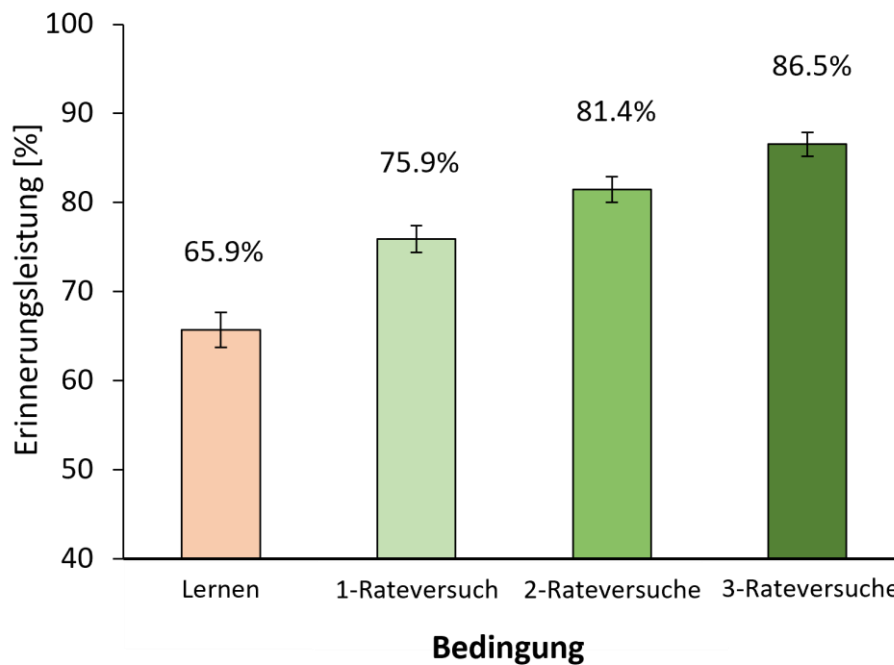
Abschließend folgte die finale Test-Phase, in der die Cue-Wörter aller 60 Wortpaare für 10 s und in zufälliger Reihenfolge präsentiert wurden und die Versuchspersonen versuchen sollten, sich an das dazugehörige Target-Wort zu erinnern (z. B. Teller - ?). Auch im finalen Test sollten die Antworten mündlich gegeben werden und es erfolgte dabei keine Rückmeldung über die Korrektheit der Antwort. Das Experiment verwendete somit ein Innersubjektdesign mit vier Bedingungen (Lern-Bedingung vs. 1-Rateversuch-Bedingung vs. 2-Rateversuche-Bedingung vs. 3-Rateversuche-Bedingung).

Ergebnisse

Initialer Vortest. Zunächst sollte betrachtet werden, inwiefern sich die Leistung im initialen Vortest in Abhängigkeit der Rateversuche verändert. Dazu wurde eine Varianzanalyse (ANOVA) mit den drei Vortestbedingungen (1-Rateversuch-Bedingung, 2-Rateversuche-Bedingung, 3-Rateversuche-Bedingung) durchgeführt, wobei sich ein signifikanter Einfluss der Bedingung auf den Anteil korrekt erratener Target-Wörter im Vortest zeigte, $F(2, 94) = 7.05$, $MSE = 22.20$, $p = .001$, $\eta^2_p = .13$. Die Versuchspersonen gaben signifikant weniger korrekte Rateversuche in der 1-Rateversuche-Bedingung ab als in der 2-Rateversuche-Bedingung (4.6 % vs. 7.6 %), $t(47) = 3.36$, $p = .002$, Cohen's $d = 0.49$, wohingegen sich die 2-Rateversuche-Bedingung und die 3-Rateversuche-Bedingung nicht signifikant voneinander unterschieden (7.6 % vs. 7.8 %), $t(47) < 1$. Da in dem vorliegenden Experiment der Einfluss falscher Rateversuche auf die spätere Erinnerungsleistung untersucht werden sollte, wurden alle Wortpaare, für die im Vortest die korrekte Lösung genannt wurde, von den nachfolgenden Analysen ausgeschlossen.

Finaler Test.

Korrektur Abruf. Abbildung 2 zeigt den Anteil korrekt erinnelter Target-Wörter im finalen Test für jede der vier Bedingungen (Lern-Bedingung vs. 1-Rateversuch-Bedingung vs. 2-Rateversuche-Bedingung vs. 3-Rateversuche-Bedingung). Eine einfaktorielle ANOVA mit den vier Bedingungen zeigte einen signifikanten Einfluss der Bedingung auf den korrekten Abruf der Wortpaare, $F(3, 141) = 112.17$, $MSE = 34.10$, $p < .001$, $\eta^2_p = .71$. Paarweise Vergleiche zeigten einen signifikanten Unterschied in der Erinnerungsleistung zwischen der Lern-Bedingung und der 1-Rateversuche-Bedingung (65.7 % vs. 75.9 %), $t(47) = 7.24$, $p < .001$, Cohen's $d = 1.05$, was einen typischen Vortesteffekt von 10.2 % widerspiegelt. Außerdem war auch der Unterschied von 5.6 % zwischen der 1-Rateversuch-Bedingung und der 2-Rateversuche-Bedingung (75.9 % vs. 81.4 %), $t(47) = 5.93$, $p < .001$, Cohen's $d = 0.86$, sowie der Unterschied von 5.1 % zwischen der 2-Rateversuche-Bedingung und der 3-Rateversuche-Bedingung signifikant (81.4 vs. 86.5 %), $t(47) = 5.04$, $p < .001$, Cohen's $d = 0.73$. Insgesamt stieg die Größe des Vortest-Effekts folglich mit zunehmender Anzahl an Rateversuchen im initialen Test von 10.2 % auf 15.8 % bzw. 20.8 %.

Abbildung 2*Ergebnisse von Experiment 1*

Anmerkung: Erinnerungsleistung im finalen Test (in %) für alle vier Bedingungen (Lernen vs. 1-Rateversuch vs. 2-Rateversuche vs. 3-Rateversuche). Fehlerbalken repräsentieren den Standardfehler des Mittelwertes.

Intrusionen. Alle fehlerhaften Antworten, die die Versuchspersonen im finalen Test abgaben, wurden als Intrusionen gezählt. Eine ANOVA mit den vier Bedingungen (Lern-Bedingung vs. 1-Rateversuch-Bedingung vs. 2-Rateversuche-Bedingung vs. 3-Rateversuche-Bedingung) zeigte dabei einen signifikanten Einfluss der Bedingung auf die Anzahl der Intrusionen, $F(3, 141) = 37.88$, $MSE = 0.95$, $p < .001$, $\eta^2_p = .45$. Geplante paarweise Vergleiche ergaben einen signifikanten Unterschied von 1.4 Intrusionen zwischen der Lern-Bedingung und der 1-Rateversuch-Bedingung (2.8 Intrusionen vs. 1.4 Intrusionen), $t(47) = 7.01$, $p < .001$, Cohen's $d = 1.02$, jedoch keine signifikanten Unterschiede in der Anzahl der Intrusionen zwischen den verschiedenen Vortest-Bedingungen, $ps > .075$.

Diskussion

Einerseits zeigten die Ergebnisse von Experiment 1 einen Erinnerungsvorteil der 1-Rateversuch-Bedingung verglichen mit der Lern-Bedingung, was den klassischen Vortesteffekt darstellt und somit die Befunde bisheriger Studien zu diesem Paradigma replizieren konnte. Andererseits liefern die vorliegenden Ergebnisse einen ersten Hinweis darauf, dass die Größe des Vortest-Effekts noch gesteigert werden kann, wenn die Versuchspersonen während der Lern-Phase nicht nur einmal, sondern mehrmals die Möglichkeit haben zu raten. So profitierten die Versuchspersonen von jedem zusätzlichen Rateversuch, was sich in einem Anstieg des Erinnerungsvorteils der jeweiligen Vortest-Bedingung verglichen mit der Lern-Bedingung zeigte und von der 1-Rateversuch-Bedingung auf die 3-Rateversuche-Bedingung nahezu zu einer Verdopplung des Vortest-Effekts führte. Auch bezüglich der Anzahl an Intrusionen profitierten die Versuchspersonen von der Bearbeitung eines Vortests, was sich in einer insgesamt geringeren Anzahl an fehlerhaften Antworten im finalen Test in den Vortest-Bedingungen verglichen mit der Lern-Bedingung zeigte, wobei dieser vorteilhafte Effekt in Experiment 1 nicht signifikant mit der Anzahl an Rateversuchen zunahm.

Das Ziel von Experiment 2 war es die Generalisierbarkeit der Befunde aus Experiment 1 und damit die Frage zu untersuchen, ob sich auch unter Verwendung eines komplexeren, pädagogisch relevanten Lernmaterials die Vorteile wiederholter Vortests zeigen. Dazu wurden in Experiment 2 Swahili-Deutsch Vokabeln als Lernmaterial verwendet und eine zusätzliche Lern-Bedingung anstelle der 2-Rateversuche-Bedingung integriert, um eine zeitlich gematchte Baseline-Bedingung für die 3-Rateversuche-Bedingung zu schaffen.

5.2. Experiment 2

Methode

Stichprobe. Wie in Experiment 1 nahmen auch an diesem Experiment 48 Versuchspersonen ($M = 24.4$ Jahre, Altersspanne = 19-34 Jahre, 39 weiblich) teil. Für alle Versuchspersonen war Deutsch die Muttersprache und zu Beginn des Experimentes wurde

mündlich das Einverständnis zur Teilnahme gegeben. Als Vergütung erhielten die Teilnehmer entweder Versuchspersonenstunden oder eine monetäre Entlohnung.

Material. Das Lernmaterial bestand aus 60 Swahili-Deutsch Vokabeln (z. B. Pombe - Bier), die aus den Normen für Swahili-Englisch Vokabeln von Nelson und Dunlosky (1994) entnommen und ins Deutsche übersetzt wurden. Die Wörter aller Vokabelpaare hatten gemäß der *Hyperspace Analogue to Language frequency scale* (Balota et al., 2007) eine Auftretenshäufigkeit von unter 100 Fällen pro eine Millionen Wörter. Erneut wurden alle Vokabelpaare in 4 Untergruppen mit jeweils 15 Items unterteilt, die gleich häufig in allen Experimentbedingungen verwendet wurden.

Design und Versuchsablauf. Der Versuchsablauf war identisch zu Experiment 1 mit dem Unterschied, dass statt der 2-Rateversuche-Bedingung eine weitere Lern-Bedingung verwendet wurde, in der die Versuchspersonen die vollständigen Wortpaare für 20 s präsentiert bekamen. Diese Änderung wurde vorgenommen, um eine Lern-Bedingung mit identischer Triallänge wie die 20 s dauernde 3-Rateversuche-Bedingung zu implementieren, um somit eine adäquate Baseline-Bedingung für den Vergleich zwischen gelernten und dreifach vorgetesteten Paaren zu schaffen. Da die Versuchspersonen in Experiment 1 bereits 10 s Zeit hatten, um sich die Wortpaare in der Lern-Bedingung einzuprägen, erscheint es unwahrscheinlich, dass sich das Verdoppeln der Lernzeit in einer zusätzlichen Lern-Bedingung erheblich auf die Erinnerungsleistung auswirkt. Um sich jedoch der Auswirkungen einer längeren Präsentation der Wortpaare in der Lern-Bedingung auf die Erinnerungsleistung völlig sicher zu sein, wurde in Experiment 2 trotzdem eine zusätzliche Baseline-Bedingung verwendet. Dementsprechend bestand das Experiment 2 aus einem 2 x 2 Design mit den beiden Innersubjekt-Faktoren ÜBUNG (Lernen vs. Vortest) und DAUER (kurz vs. lang). Die Hälfte der Vokabelpaare wurde vollständig, d. h. das Swahili-Cue-Wort zusammen mit der deutschen Übersetzung als Target-Wort (z. B. Tumbili - Affe), präsentiert und konnte somit direkt gelernt werden (Lern-Bedingung). Dabei wurde variiert, ob die Vokabelpaare für 10 s (Lern-Bedingung kurz) oder für 20 s (Lern-Bedingung lang) auf dem Bildschirm sichtbar blieben. Für die andere Hälfte der Wortpaare wurde zunächst nur das Cue-Wort präsentiert und die Versuchspersonen sollten versuchen, das dazugehörige Target-Wort zu erraten (z. B. Pombe - ?), bevor das vollständige Vokabelpaar zu sehen war (Vortest-Bedingung). Dabei sollten die Versuchspersonen für 15 der Vokabelpaare in der Vortest-

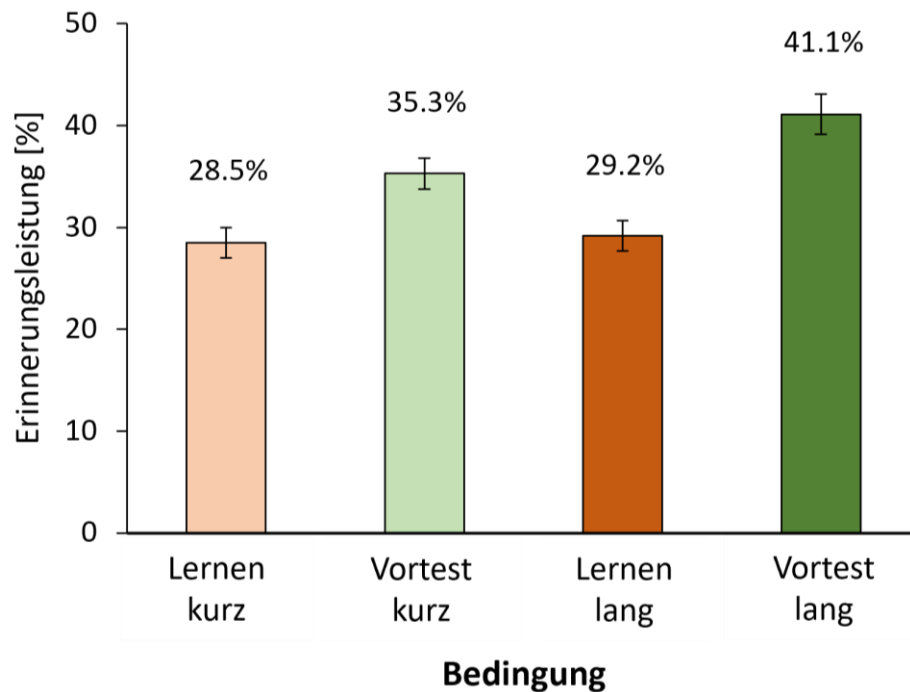
Bedingung genau einen Rateversuch (Vortest-Bedingung kurz), für die anderen 15 Vokabelpaare immer genau drei Rateversuche (Vortest-Bedingung lang) abgeben.

Ergebnisse

Initialer Vortest. Im initialen Vortest wurde in keiner der beiden Vortest-Bedingungen ein Target-Wort korrekt erraten. Somit wurden alle Items in die nachfolgenden Analysen miteingeschlossen.

Finaler Test.

Korrektur Abruf. Abbildung 3 zeigt den prozentualen Anteil korrekt erinnerter Target-Wörter im finalen Test für alle vier Bedingungen (Lern-Bedingung niedrig vs. Vortest-Bedingung niedrig vs. Lern-Bedingung hoch vs. Vortest-Bedingung hoch). Eine 2 x 2 ANOVA mit den Faktoren ÜBUNG (Lernen vs. Vortest) und DAUER (kurz vs. lang) zeigte sowohl einen Haupteffekt ÜBUNG, $F(1, 47) = 62.99$, $MSE = 40.30$, $p < .001$, $\eta^2_p = .57$, mit einer besseren Erinnerungsleistung für vorgetestete Vokabelpaare verglichen mit ausschließlich gelernten Vokabelpaaren (38.2 % vs. 28.8 %) als auch einen Haupteffekt DAUER, $F(1, 47) = 13.78$, $MSE = 40.30$, $p = .001$, $\eta^2_p = .23$, mit einer besseren Erinnerungsleistung für Wortpaare in der Bedingung mit langer Dauer verglichen mit der Bedingung mit kurzer Dauer (35.1 % vs. 31.9 %). Vor allem aber zeigte sich auch eine signifikante Interaktion zwischen den beiden Faktoren $F(1, 47) = 7.86$, $MSE = 40.30$, $p = .007$, $\eta^2_p = .14$. Geplante paarweise Vergleiche ergaben, dass Vokabelpaare in der Vortestbedingung, verglichen mit der entsprechenden Lern-(Baseline-)Bedingung, besser wiedererinnert wurden, wenn die Dauer des Lern-Trials kurz, d. h. 10 s, war, $t(47) = 4.83$, $p < .001$, Cohen's $d = 0.70$, und dieser Erinnerungsvorteil sogar noch größer ausfiel, wenn die Dauer des Lern-Trials lang, d. h. 20 s, war, $t(47) = 7.58$, $p < .001$, Cohen's $d = 1.10$. Folglich stieg auch die Größe des Vortest-Effekts von 6.8 % (35.3 % vs. 28.5 %) in der Bedingung mit kurzer Dauer des Lern-Trials auf 11.9 % (41.1 % vs. 29.2 %) in der Bedingung mit langer Dauer des Lern-Trials.

Abbildung 3*Ergebnisse von Experiment 2*

Anmerkung: Erinnerungsleistung im finalen Test (in %) in Abhängigkeit der beiden Faktoren Bedingung (Lernen vs. Vortest) und Dauer (kurz vs. lang). Fehlerbalken repräsentieren den Standardfehler des Mittelwertes.

Intrusionen. Eine 2 x 2 ANOVA der beiden Faktoren ÜBUNG (Lernen vs. Vortest) und DAUER (kurz vs. lang) offenbarte einen Haupteffekt ÜBUNG $F(1, 47) = 67.96$, $MSE = 1.33$, $p < .001$, $\eta^2_p = .59$, mit einer niedrigeren Anzahl an Intrusionen in der Vortest-Bedingung verglichen mit der Lern-Bedingung (1.6 Intrusionen vs. 2.8 Intrusionen) und einen Haupteffekt DAUER, $F(1, 47) = 4.53$, $MSE = 1.33$, $p = .039$, $\eta^2_p = .09$, mit einer niedrigeren Anzahl an Intrusionen in der Bedingung mit langer Dauer des Lern-Trials verglichen mit der Bedingung mit kurzer Dauer des Lern-Trials (2.1 Intrusionen vs. 2.3 Intrusionen). Es zeigte sich allerdings keine signifikante Interaktion zwischen den beiden Faktoren, $F(1, 47) = 2.26$, $MSE = 1.33$, $p = .14$, $\eta^2_p = .05$.

Diskussion

Mit den vorliegenden Befunden ließen sich die Ergebnismuster aus Experiment 1 replizieren. Erneut zeigte sich in der Bedingung mit kurzer Dauer des Lern-Trials beim

Vergleich der Lern-Bedingung mit der Vortest-Bedingung ein typischer Vortesteffekt. Dieser Vortesteffekt nahm erneut mit der Anzahl der Rateversuche zu, denn wenn Versuchspersonen im initialen Vortest drei Rateversuche generieren sollten, fiel der Erinnerungsvorteil der Vortest-Bedingung gegenüber der Lern-Bedingung fast doppelt so groß aus, als wenn die Versuchspersonen nur einen Rateversuch abgeben sollten. Die Ergebnisse von Experiment 2 liefern also einerseits weitere Hinweise darauf, dass die Anzahl der Vortests einen positiven Einfluss auf die nachfolgende Erinnerungsleistung der vorgetesteten Items hat, andererseits erweitern und generalisieren die Ergebnisse die Befunde von Experiment 1 dahingehend, dass sich dieser Einfluss auch bei pädagogisch relevantem Lernmaterial wie Vokabeln zeigt. Selbst wenn der Vortest wie im vorliegenden Experiment so schwierig ist, dass kein einziges Target-Wort korrekt erraten wird, scheinen die Versuchspersonen also stärker zu profitieren, wenn sie mehrere (falsche) Rateversuche abgeben sollen.

Das in den Experimenten 1 und 2 gewählte Vorgehen führte zu einer natürlichen Störvariable, denn in beiden Experimenten stieg mit der Anzahl an Rateversuchen auch die Dauer des Lern-Trials in der jeweiligen Bedingung. Dadurch bleibt es zunächst noch unklar, ob es die Anzahl der Rateversuche an sich ist, die für die Vergrößerung des Vortest-Effekts verantwortlich war oder ob es die längere Präsentationszeit des Lernmaterials war, die zu den beobachteten Effekten führte. Um sich dieser Fragestellung zu widmen, wurde Experiment 3 durchgeführt. Das Ziel dieses Experiments war es, zu untersuchen, ob sich der in der Vortest-Bedingung gefundene Erinnerungsvorteil von wiederholten Rateversuchen verglichen mit nur einem Rateversuch auch dann noch zeigt, wenn die Trial-Länge dieser beiden Bedingungen konstant gehalten wird. Wenn tatsächlich die Anzahl an Rateversuchen entscheidend für den Anstieg des Vortest-Effekts in den Experimenten 1 und 2 war, dann sollte sich auch in Experiment 3, trotz identischer Trial-Länge in den beiden Bedingungen, noch immer ein größerer Vortest-Effekt für die 3-Rateversuche-Bedingung verglichen mit der 1-Rateversuch-Bedingung zeigen. War es hingegen die längere Zeitspanne, die ursächlich für den Anstieg des Vortest-Effekts in den vorherigen beiden Experimenten war, so sollte der Vortest-Effekt im nachfolgenden Experiment 3 nicht von der Anzahl an Rateversuchen beeinflusst werden und sich somit keine Zunahme des Vortest-Effekts mehr zeigen.

5.3. Experiment 3

Methode

Stichprobe. Der Poweranalyse aus Experiment 1 folgend nahmen auch an diesem Experiment 48 Versuchspersonen ($M = 24.9$ Jahre, Altersspanne = 19-32 Jahre, 38 weiblich) teil. Erneut sprachen alle Versuchspersonen Deutsch als Muttersprache und gaben zu Beginn des Experimentes ihr mündliches Einverständnis. Als Vergütung wurden wahlweise Versuchspersonenstunden oder eine monetäre Entlohnung vergeben.

Material. Als Lernmaterial wurden dieselben Wortpaare wie in Experiment 1 verwendet und die Zuordnung zu den Experimentalbedingungen erfolgte wie in den Experimenten 1 und 2.

Design und Versuchsablauf. Mit Ausnahme von zwei Details war der Aufbau und Ablauf von Experiment 3 identisch zu Experiment 1. Einerseits wurde ein Innersubjektdesign mit drei (statt vier) Bedingungen verwendet: einer Lern-Bedingung, einer 1-Rateversuch-Bedingung und einer 3-Rateversuche-Bedingung. Dabei wurden den Versuchspersonen in der Lern-Bedingung die vollständigen Wortpaare für 14 s präsentiert. In der 1-Rateversuch-Bedingung wurde zunächst nur das Cue-Wort präsentiert und die Versuchspersonen hatten 9 s Zeit, um einen Rateversuch für das korrekte Target-Wort zu generieren, bevor im Anschluss das intakte Wortpaar für 5 s zu sehen war. Auch in der 3-Rateversuche-Bedingung wurde für 9 s zunächst nur das Cue-Wort präsentiert. Die Versuchspersonen sollten in dieser Zeit allerdings versuchen, drei unterschiedliche Rateversuche für das dazugehörige Target-Wort abzugeben, bevor das intakte Wortpaar für 5 s gezeigt wurde. Sowohl in der 1-Rateversuch-Bedingung als auch in der 3-Rateversuche-Bedingung dauerte ein Trial folglich insgesamt 14 s, identisch zu der Dauer eines Lern-Trials.

Andererseits wurden in Experiment 3 die 1-Rateversuch-Bedingung und die 3-Rateversuche-Bedingung (anders als in Experiment 1) in der Lern-Phase in getrennten Blöcken verwendet. Für die Hälfte der Versuchspersonen begann die Lern-Phase mit einem Block, in dem für 15 Wortpaare zunächst für 9 s nur das Cue-Wort gezeigt und genau ein Rateversuch abgegeben werden sollte, bevor das intakte Wortpaar im Anschluss für 5 s zu

sehen war (1-Rateversuch-Bedingung). Die anderen 15 Wortpaare in diesem Block wurden vollständig für 14 s präsentiert (Lern-Bedingung), wobei die Reihenfolge der Präsentation der 30 Wortpaare des ersten Blocks zufällig war. Direkt im Anschluss folgte ein zweiter Block, in dem für 15 Wortpaare erneut für 9 s nur das Cue-Wort zu sehen war, jedoch sollten die Versuchspersonen hier jeweils innerhalb von 3 s einen neuen Rateversuch, also insgesamt 3 Rateversuche innerhalb der 9 s generieren (3-Rateversuche-Bedingung), bevor das vollständige Wortpaar für 5 s präsentiert wurde. Wie im ersten Block wurden auch in diesem Block 15 weitere Wortpaare verwendet, die direkt für 14 s vollständig präsentiert wurden (Lern-Bedingung) und zusammen mit den anderen Wortpaaren in zufälliger Reihenfolge gezeigt wurden. Für die andere Hälfte der Versuchspersonen war die Zuordnung der Experimentalbedingungen in den ersten und zweiten Block genau umgekehrt. Zu Beginn eines jeden Blocks wurden die Versuchspersonen informiert, wie viele Antworten in den Rate-Trials im Anschluss abgegeben werden sollten.

Ergebnisse

Initialer Vortest. Ein paarweiser Vergleich zeigte, dass Versuchspersonen in der 1-Rateversuch-Bedingung signifikant weniger korrekte Rateversuche produzierten als in der 3-Rateversuche-Bedingung (4.4 % vs. 9.7 %), $t(47) = 5.55$, $p < .001$, Cohen's $d = 0.80$. Da erneut der Effekt falscher Rateversuche auf die nachfolgende Gedächtnisleistung von Interesse war, wurden alle Wortpaare, für die im Vortest die korrekte Lösung erraten wurde, von den nachfolgenden Analysen ausgeschlossen.

Ein direkter Vergleich bezüglich der Anzahl korrekter erratener Target-Wörter im initialen Vortest zwischen Experiment 1 und Experiment 3 zeigte weder für die 1-Rateversuch-Bedingung (4.6 % vs. 4.4 %), $t(94) < 1$, Cohen's $d = 0.03$, noch für die 3-Rateversuche-Bedingung (7.8 % vs. 9.7 %), $t(94) = 1.47$, $p = .14$, Cohen's $d = 0.30$, einen signifikanten Unterschied. Dies legt nahe, dass die verfügbare Zeit für einen einzelnen Rateversuch keine entscheidende Rolle für die Anzahl korrekter Rateversuche spielt.

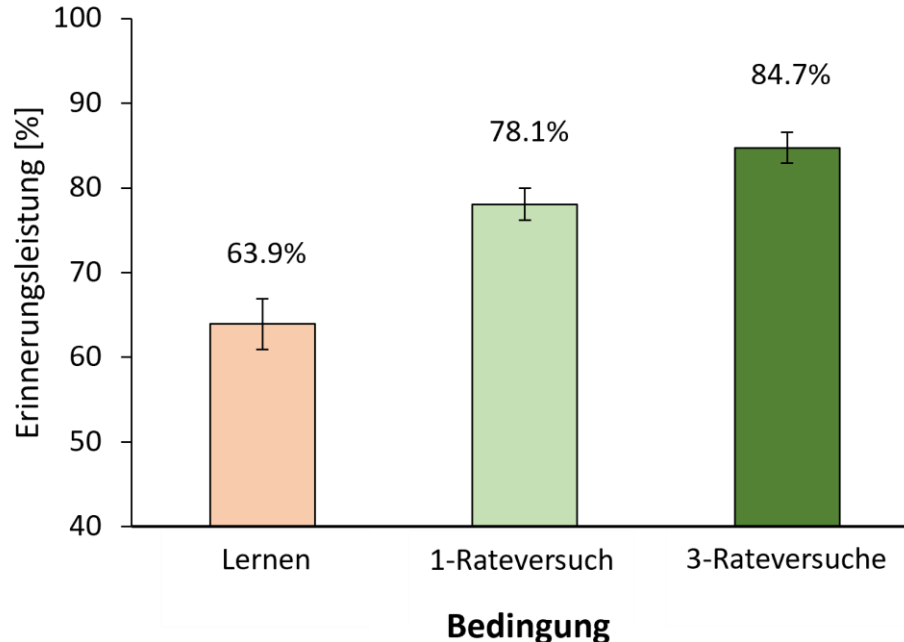
Finaler Test.

Korrektur Abruf. Da sich die durchschnittliche Erinnerungsleistung der Lern-Bedingung der beiden Blöcke nicht signifikant voneinander unterschied, $t(47) < 1$, Cohen's $d = 0.01$, wurden für die nachfolgenden Analysen alle Trials der Lern-Bedingung

zusammengefasst. In Abbildung 4 ist der prozentuale Anteil der korrekt erinnerten Target-Items im finalen Test für jede der drei Bedingungen (Lern-Bedingung, 1-Rateversuch-Bedingung, 3-Rateversuche-Bedingung) dargestellt. Eine einfaktorielle ANOVA der drei Bedingungen zeigte dabei einen signifikanten Effekt der Bedingung auf die Erinnerungsleistung, $F(2, 94) = 44.10$, $MSE = 123.08$, $p < .001$, $\eta^2_p = .48$. Geplante paarweise Vergleiche legten offen, dass der Unterschied von 14.2 % zwischen der Lern-Bedingung und der 1-Rateversuchbedingung (63.9 % vs. 78.1 %) signifikant war, $t(47) = 5.53$, $p < .001$, Cohen's $d = 0.80$, was den typischen Vortest-Effekt darstellte. Außerdem war auch der Unterschied von 6.7 % zwischen der 1-Rateversuch-Bedingung und der 3-Rateversuche-Bedingung (84.8 % vs. 78.1 %) signifikant, $t(47) = 3.82$, $p < .001$, Cohen's $d = 0.55$, womit verglichen mit der Lern-Bedingung die Größe des Vortest-Effekts von 14.2 % in der 1-Rateversuch-Bedingung auf 20.8 % in der 3-Rateversuche-Bedingung stieg.

Abbildung 4

Erinnerungsleistung in Experiment 3



Anmerkung: Erinnerungsleistung im finalen Test (in %) für alle drei Bedingung (Lernen vs. 1-Rateversuch vs. 3-Rateversuche). Fehlerbalken repräsentieren den Standardfehler des Mittelwertes.

Intrusionen. Bei der Analyse der fehlerhaften Antworten im finalen Test zeigte eine einfaktorielle ANOVA mit den drei Bedingungen (Lern-Bedingung, 1-Rateversuch-Bedingung, 3-Rateversuche-Bedingung) einen signifikanten Einfluss der Bedingung auf die Anzahl an Intrusionen, $F(2, 94) = 28.80$, $MSE = 1.72$, $p < .001$, $\eta^2_p = .38$. Geplante paarweise Vergleiche offenbarten dabei, dass in der Lern-Bedingung signifikant mehr Intrusionen produziert wurden als in der 1-Rateversuch-Bedingung (2.9 Intrusionen vs. 1.5 Intrusionen), $t(47) = 4.37$, $p < .001$, Cohen's $d = 0.63$ und zudem in der 3-Rateversuche-Bedingung signifikant weniger Intrusionen produziert wurden als in der 1-Rateversuche-Bedingung (0.9 Intrusionen vs. 1.5 Intrusionen), $t(47) = 3.42$, $p = .001$, Cohen's $d = 0.49$. Somit nahm auch dieser vorteilhafte Effekt von Vortests verglichen mit der Lern-Bedingung von 1.4 Intrusionen in der 1-Rateversuch-Bedingung auf 2.0 Intrusionen in der 3-Rateversuche-Bedingung zu.

Diskussion

Wie bereits in den Experimenten 1 und 2 zeigte sich erneut ein klassischer Vortest-Effekt in der 1-Rateversuch-Bedingung verglichen mit der Lern-Bedingung, der in seiner Größe noch weiter zunahm, wenn die Versuchspersonen statt einmal sogar dreimal das zugehörige Target-Wort erraten sollten (3-Rateversuche-Bedingung). Die Replikation des Befundmusters der beiden vorangegangenen Experimente konnte in Experiment 3 gefunden werden, obwohl die Dauer eines Trials innerhalb der Vortest-Bedingungen unabhängig von der Anzahl an Rateversuchen konstant bei 14 s gehalten wurde. Diese Ergebnisse lassen es unwahrscheinlich erscheinen, dass die in den Experimenten 1 und 2 beobachtete Zunahme des Vortest-Effekts in Abhängigkeit der steigenden Anzahl an Rateversuchen auf die verlängerte Dauer eines Trials zurückgeführt werden kann. Vielmehr scheint es die Anzahl der Rateversuche an sich zu sein, die sich positiv auf die Größe des Vortest-Effekts auswirkt.

Mit dem abschließend durchgeführten Experiment 4 sollten zwei Hauptziele verfolgt werden. Einerseits sollte untersucht werden, ob wiederholte Rateversuche auch dann noch zu einer Verbesserung der Gedächtnisleistung führen, wenn ein komplexes, pädagogisch relevantes Textmaterial als Lerninhalt verwendet wird. Andererseits sollte mit Experiment 4 die Frage untersucht werden, ob wiederholte Vortests auch zu einer verbesserten Erinnerung an zuvor gelernte, aber nicht abgetestete Information beitragen können. Vor allem die Frage, ob Vortests auch zu einem Transfereffekt auf ungetestetes Material führen können, ist aus

einer angewandten Perspektive entscheidend, da es Lehrpersonen in realen Lernsituationen typischerweise vermeiden, während des Unterrichts spätere Prüfungsfragen zu stellen (z. B. Wixted et al., 2014). Während die bisherige Studienlage zu dieser Fragestellung die Vermutung nahelegt, dass sich zumindest in der Vortest-Bedingung mit einem Rateversuch typischerweise kein Transfereffekt auf ungetestetes Material zeigen sollte (z. B. James & Storm, 2019; Richland et al., 2009), muss jedoch darauf hingewiesen werden, dass in diesen Studien alle Vortest-Fragen in einem separaten Block beantwortet wurden, bevor das Material gelernt wurde.

5.4. Experiment 4

Method

Stichprobe. Basierend auf der Poweranalyse aus Experiment 1 nahmen an diesem Experiment 54 Versuchspersonen ($M = 25.6$ Jahre, Altersspanne = 20-38 Jahre, 39 weiblich) teil. Um eine Ausbalancierung des Materials und der Bedingungen zu gewährleisten, nahmen an Experiment 4 sechs Versuchspersonen mehr als an den vorherigen Experimenten teil. Für alle Versuchspersonen stellte Deutsch die Muttersprache dar und alle Versuchspersonen gaben ihr mündliches Einverständnis zur Teilnahme. Als Entlohnung wurden entweder eine monetäre Vergütung oder Versuchspersonenstunden vergeben.

Material. Das Lernmaterial in Experiment 4 bestand aus drei lexikonartig aufbereiteten Texten über das Land Brasilien (266 Wörter), den ehemaligen US-Präsidenten Woodrow Wilson (233 Wörter) und Tapire (252 Wörter). Dazu wurden Informationen aus Wikipedia über die jeweiligen Themen verwendet (<https://de.wikipedia.org/wiki/Brasilien>; https://de.wikipedia.org/wiki/Woodrow_Wilson; <https://de.wikipedia.org/wiki/Tapire>). Jeder dieser Texte bestand dabei aus 12 Sätzen und jeder Text wurde gleich häufig in jeder der drei Bedingungen (Lern-Bedingung, 1-Rateversuch-Bedingung, 3-Rateversuche-Bedingung) verwendet. Für jeden dieser Sätze wurde eine Frage für den finalen Test erstellt, die sich auf eine im Satz enthaltenen Information bezog. Alle Fragen wurden dabei als offene Fragen im Kurzantwort-Format konzipiert (z. B. „Was war der Vorname von Woodrow Wilsons‘ Mutter? Wann erlangte Brasilien seine Unabhängigkeit von Portugal? Wie viele

lebende Arten der Gattung Tapire gibt es noch?“). Diese 12 Fragen pro Text wurden in zwei Untergruppen bestehend aus jeweils sechs Fragen (Untergruppe A und Untergruppe B) unterteilt. Für die Hälfte der Versuchspersonen wurde die für die Beantwortung der sechs Fragen aus Untergruppe A notwendige Information im initialen Vortest als Satzlücke präsentiert und die Information für die Fragen der Untergruppe B war vollständig sichtbar. Diese Zuordnung der Untergruppen auf die Bedingungen war für die andere Hälfte der Versuchspersonen genau umgekehrt. Im finalen Test sollten immer alle 12 Fragen der beiden Untergruppen beantwortet werden.

Design und Versuchsablauf. Wie die bisherigen Experimente 1-3 wurde auch Experiment 4 via Zoom durchgeführt. Dabei wurde ein Innersubjekt-Design mit den drei Bedingungen: Lern-Bedingung, 1-Rateversuch-Bedingung und 3-Rateversuche-Bedingung verwendet. In allen drei Bedingungen durchliefen die Versuchspersonen immer zunächst eine Lern-Phase, in der einer der drei Texte gelernt werden sollte, bevor der Inhalt dieses Textes nach einer kurzen Distraction in einem finalen Test abgefragt wurde, d. h. nach der Lern-Phase von Text 1 folgte anschließend der finale Test von Text 1, bevor anschließend die Texte 2 und 3 erst gelernt und dann abgefragt wurden. Es gab am Ende keinen kumulativen Test über den Inhalt aller drei Texte. Die Lern-Phase der Texte unterschied sich dabei in Abhängigkeit der Bedingung. In der Lern-Phase der Lern-Bedingung wurden den Versuchspersonen der Reihe nach alle 12 Sätze für jeweils 20 s vollständig auf dem Bildschirm präsentiert (z. B. „Kulturell ist Brasilien durch Musikformen wie den Samba und den Bossa Nova geprägt.“). In der Lern-Phase der 1-Rateversuch-Bedingung wurden ebenfalls alle 12 Sätze der Reihe nach präsentiert, jedoch war nur die Hälfte der Sätze von Beginn an vollständig sichtbar und verblieb für 20 s auf dem Bildschirm. Bei der anderen Hälfte der Sätze war zunächst eine Lücke im Satz sichtbar und die Versuchspersonen hatten 10 s Zeit, den Lückensatz zu lesen, bevor sie durch einen Ton, der 0.5 s dauerte, dazu aufgefordert wurden, einen Rateversuch für die fehlende Information zu generieren. Für diesen Rateversuch hatten die Versuchspersonen 5 s Zeit und als Hilfestellung wurde hinter der Lücke in Klammern vermerkt, um welche Art von Antwort es sich dabei handelt (z. B. „Brasilien erlangte ____ (Zahl) seine Unabhängigkeit von Portugal.“), bevor anschließend der Satz inklusive der korrekten Antwort für weitere 5 s zu sehen war (z. B. „Brasilien erlangte 1822 seine Unabhängigkeit von Portugal.“). Die Lern-Phase der 3-Rateversuche-

Bedingung war identisch zu der Lern-Phase der 1-Rateversuch-Bedingung mit dem einzigen Unterschied, dass die Versuchspersonen für jede Lücke drei unterschiedliche Rateversuche generieren sollten, für die sie jeweils 5 s Zeit hatten. Dabei wurde jeder Rateversuch durch einen Ton angekündigt. In den beiden Vortest-Bedingungen wurden vollständige Sätze und Lückensätze jeweils abwechselnd präsentiert.

Nach der Lern-Phase des Textes folgte eine 5-minütige Distraktor-Phase in der die Versuchspersonen zunächst für 1 min in 7er-Schritten rückwärts zählen sollten, bevor für weitere 4 min in einem Online-Spiel die Unterschiede zwischen zwei Bildern gesucht werden sollten (<https://www.suchbilder.com/fehlerbilder/>). Anschließend wurde der finale Test über den vorangegangenen Text bearbeitet. Bei diesem finalen Test wurden den Versuchspersonen 12 Fragen über den Inhalt des Textes gestellt, wobei jede Frage den Inhalt eines der 12 präsentierten Sätze abprüfte (z. B. „Durch welche Musikform neben dem Samba ist Brasilien kulturell geprägt?“). Die Versuchspersonen hatten dabei 8 s Zeit mündlich ihre Antwort zu geben und es erfolgte kein Feedback über die Korrektheit der Antwort. Nach dem finalen Test wurden die Versuchspersonen aufgefordert 3 min Tetris zu spielen (<https://www.geo.de/geolino/spiele/13349-rtklonlinespiel-tetris>), bevor anschließend die Lern-Phase des nächsten Textes stattfand. Jeder Text wurde dabei gleich häufig als Text 1, Text 2 oder Text 3 verwendet und auch jede der drei Bedingungen wurde gleich häufig in der Lern-Phase von Text 1, Text 2 und Text 3 verwendet.

Ergebnisse

Initialer Vortest. Ein paarweiser Vergleich zeigte, dass Versuchspersonen in der 3-Rateversuch-Bedingung numerisch zwar mehr korrekte Rateversuche produzierten als in der 1-Rateversuch-Bedingung (5.2 % vs. 3.1 %), allerdings war dieser Unterschied nicht signifikant, $t(53) = 1.63$, $p = .11$, Cohen's $d = 0.22$. Da erneut der Effekt falscher Rateversuche auf die nachfolgende Gedächtnisleistung von Interesse war, wurden alle Wortpaare, für die im Vortest die korrekte Lösung erraten wurde, von den nachfolgenden Analysen ausgeschlossen.

Finaler Test.

Korrektur Abruf. Abbildung 5 zeigt den prozentualen Anteil an korrekt beantworteten Fragen im finalen Test für die Lern-Bedingung und für die 1-Rateversuch-Bedingung und

die 3-Rateversuche-Bedingung jeweils unterteilt in vorgetestete und ungetestete Sätze. Im ersten Schritt wurde zum Vergleich der Erinnerungsleistung im finalen Test zwischen der Lern-Bedingung und dem vorgetesteten Material der 1-Rateversuch-Bedingung und der 3-Rateversuche-Bedingung eine einfaktorielle ANOVA durchgeführt, die einen signifikanten Einfluss der Bedingung auf die korrekte Erinnerungsleistung zeigte, $F(2,106) = 42.21$, $MSE = 227.94$, $p < .001$, $\eta^2_p = .44$. Geplante paarweise Vergleiche offenbarten dabei eine signifikant bessere Erinnerungsleistung in der 1-Rateversuch-Bedingung als in der Lern-Bedingung (78.3 % vs. 59.4 %), $t(53) = 5.94$, $p < .001$, Cohen's $d = 0.81$, was einen klassischen Vortest-Effekt von 18.9 % darstellt. Zudem zeigte sich ein signifikanter Unterschied von 6.9 % zwischen der 1-Rateversuch-Bedingung und der 3-Rateversuche-Bedingung (85.2 % vs. 78.3 %), $t(53) = 2.54$, $p = .014$, Cohen's $d = 0.35$, wodurch die Größe des Vortest-Effekts für getestetes Material von 18.9 % in der 1-Rateversuch-Bedingung auf 25.8 % in der 3-Rateversuche-Bedingung anstieg.

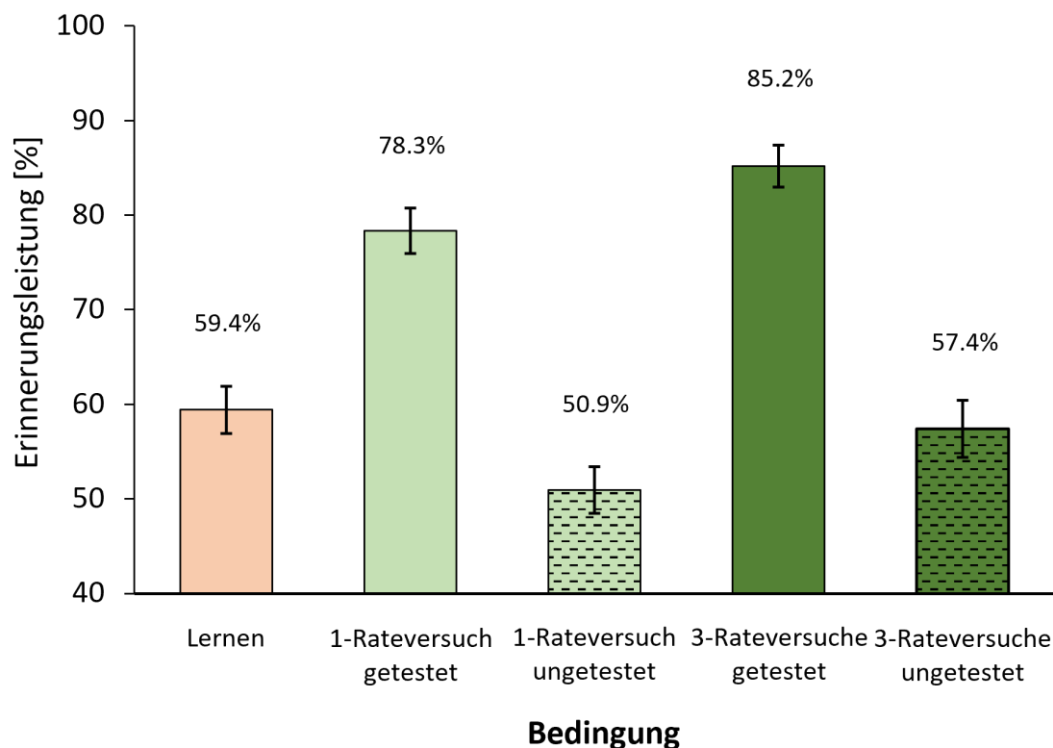
In einem zweiten Analyse-Schritt wurde erneut eine einfaktorielle ANOVA zum Vergleich der Lern-Bedingung, diesmal mit dem jeweils ungetesteten Material der 1-Rateversuch-Bedingung und der 3-Rateversuche-Bedingung, durchgeführt. Dabei zeigte sich ein signifikanter Einfluss der Bedingung, $F(2,106) = 3.09$, $MSE = 343.77$, $p = .05$, $\eta^2_p = .06$. Geplante paarweise Vergleiche offenbarten einen signifikanten Unterschied von 8.5 % zwischen der Lern-Bedingung und der 1-Rateversuch-Bedingung (59.4 % vs. 50.9 %), $t(53) = 2.44$, $p = .018$, Cohen's $d = 0.33$, wohingegen weder der Unterschied von 6.5 % zwischen der 1-Rateversuch-Bedingung und der 3-Rateversuche-Bedingung (57.4 % vs. 50.9 %) noch der Unterschied von 2.0 % zwischen der Lern-Bedingung und der 3-Rateversuche-Bedingung (59.4 % vs. 57.4 %) signifikant waren, alle $ps > .10$.

Im dritten und letzten Analyse-Schritt der korrekten Abrufleistung wurde eine 2×2 ANOVA mit den Innersubjekt-Faktoren ÜBUNG (1-Rateversuch-Bedingung vs. 3-Rateversuche-Bedingung) und MATERIAL (vorgetestet vs. ungetestet) durchgeführt, um zu untersuchen, ob sich der Anstieg der Erinnerungsleistung von der 1-Rateversuch-Bedingung auf die 3-Rateversuche-Bedingung für vorgetestetes und ungetestetes Material unterschied. Die ANOVA zeigte dabei Haupteffekte der Faktoren ÜBUNG, $F(1,53) = 8.91$, $MSE = 337.70$, $p = .004$, $\eta^2_p = .14$, und MATERIAL, $F(1,53) = 125.89$, $MSE = 337.70$, $p < .001$, $\eta^2_p = .70$, mit einer insgesamt besseren Erinnerungsleistung nach drei Rateversuchen

verglichen mit einem einzigen Rateversuch (71.3 % vs. 64.4 %) und für vorgetestetes Material verglichen mit ungetestetem Material (81.8 % vs. 54.2 %). Es zeigte sich keine Interaktion zwischen den beiden Faktoren $F(1,53) < 1$, was die Schlussfolgerung nahelegt, dass drei Rateversuche im Vergleich zu nur einem Rateversuch für vorgetestetes und ungetestetes Material zu einem vergleichbar großen Erinnerungsvorteil führten.

Abbildung 5

Ergebnisse von Experiment 4



Anmerkung: Erinnerungsleistung im finalen Test (in %) für die Lern-Bedingung und jeweils getrennt für getestetes und ungetestetes Material für die 1-Rateversuch-Bedingung und die 3-Rateversuche-Bedingung. Fehlerbalken repräsentieren den Standardfehler des Mittelwertes.

Intrusionen. Auch bei der Betrachtung der Intrusionen wurde die Auswertung der Ergebnisse in zwei Schritte unterteilt. Im ersten Schritt wurde zum Vergleich der Anzahl an Intrusionen zwischen der Lern-Bedingung und dem vorgetesteten Material der 1-

Rateversuch-Bedingung und der 3-Rateversuche-Bedingung eine einfaktorielle ANOVA durchgeführt, die einen signifikanten Einfluss der Bedingung auf die Anzahl an Intrusionen zeigte, $F(2,106) = 82.58$, $MSE = 1.07$, $p < .001$, $\eta^2_p = .61$. Geplante paarweise Vergleiche offenbarten dabei, dass der Unterschied von 2.0 Intrusionen zwischen der Lern-Bedingung und der 1-Rateversuch-Bedingung (2.8 Intrusionen vs. 0.8 Intrusionen) ebenso signifikant wurde, $t(53) = 8.46$, $p < .001$, Cohen's $d = 1.15$, wie der Unterschied von 0.4 Intrusionen zwischen der 1-Rateversuch-Bedingung und der 3-Rateversuche-Bedingung (0.8 Intrusionen vs. 0.4 Intrusionen), $t(53) = 2.76$, $p = .008$, Cohen's $d = 0.38$.

In einem zweiten Schritt wurde eine weitere einfaktorielle ANOVA mit der Lern-Bedingung und dem jeweils ungetesteten Material der 1-Rateversuch-Bedingung und der 3-Rateversuche-Bedingung durchgeführt, die erneut einen signifikanten Einfluss der Bedingung auf die Anzahl an Intrusionen im finalen Test zeigte, $F(2,106) = 28.87$, $MSE = 1.16$, $p < .001$, $\eta^2_p = .35$. Geplante paarweise Vergleiche offenbarten hier einen signifikanten Unterschied von 1.2 Intrusionen zwischen der Lern-Bedingung und der 1-Rateversuch-Bedingung (2.8 Intrusionen vs. 1.6 Intrusionen), $t(53) = 4.74$, $p < .001$, Cohen's $d = 0.65$, aber keinen signifikanten Unterschied zwischen der 1-Rateversuch-Bedingung und der 3-Rateversuche-Bedingung (1.6 Intrusionen vs. 1.3 Intrusionen), $t(53) = 1.78$, $p = .08$, Cohen's $d = 0.24$.

Diskussion

Die Ergebnisse von Experiment 4 zeigen einen signifikanten Vorteil von Vortests auf die spätere Erinnerungsleistung dieses vorgetesteten Materials nach einem Rateversuch verglichen mit der Lern-Bedingung, was eine Replikation des typischen Vortest-Effekts unter Verwendung von Prosa-Texten als Lernmaterial darstellt (James & Storm, 2019; Richland et al., 2009). Vor allem aber nahm die Größe des Vortest-Effekts erneut zu, wenn statt nur eines Rateversuchs drei Rateversuche generiert werden sollten, was die Befunde der Experimente 1-3 einerseits repliziert und andererseits von der Verwendung von Wortpaaren auch für die Verwendung von komplexerem Lernmaterial verallgemeinert. Zusätzlich legen die vorliegenden Ergebnisse nahe, dass Vortests keinen positiven Effekt auf die spätere Erinnerungsleistung ungetesteter Information haben, unabhängig davon, ob im Vortest ein Rateversuch oder drei Rateversuche generiert werden sollten. Dieser Befund lässt sich gut in die bisherige Literatur einordnen, die keinen vorteilhaften Einfluss eines Rateversuchs auf

die Erinnerungsleistung von gelerntem aber ungetestetem Material zeigt (James & Storm, 2019; Richland et al., 2009). Der Abruf der ungetesteten Information der 1-Rateversuch-Bedingung war sogar signifikant schlechter als in der Lern-Bedingung. Allerdings konnte auch für ungetestetes Material ein ähnlich großer Anstieg der Erinnerungsleistung von der 1-Rateversuch-Bedingung auf die 3-Rateversuche-Bedingung gefunden werden wie für vorgetestetes Material, was zu einer vergleichbaren Erinnerungsleistung für das ungetestete Material der 3-Rateversuche-Bedingung und der Lern-Bedingung führte.

Zusätzliche Analysen

Da die separat für jedes Experiment durchgeführten Analysen der gesamten Intrusionen nur bedingt Rückschlüsse über den Einfluss von (wiederholten) Vortests auf die Anzahl fehlerhafter Antworten im finalen Test zulassen, wurde abschließend eine Gesamtanalyse der Intrusionen über alle Experimente hinweg durchgeführt. Für diese Analyse wurde die Intrusionsrate – ermittelt aus der Anzahl an Intrusionen einer Bedingung geteilt durch die Gesamtanzahl an Items in der jeweiligen Bedingung – für die Lern-Bedingung, die 1-Rateversuch-Bedingung und die 3-Rateversuche-Bedingung über alle vier Experimente hinweg zusammengefasst und gemittelt. Eine einfaktorielle ANOVA der drei Bedingungen ergab einen signifikanten Einfluss der Bedingung auf die Intrusionsrate, $F(2,394) = 81.81$, $MSE = 75.39$, $p < .001$, $\eta^2_p = .29$. Geplante paarweise Vergleiche ergaben, dass der Unterschied von 7.8 % zwischen der Lern-Bedingung und der 1-Rateversuch-Bedingung (18.8 % vs. 11.0 %) ebenso signifikant war, $t(197) = 7.58$, $p < .001$, Cohen's $d = 0.54$, wie der Unterschied von 3.8 % zwischen der 1-Rateversuch-Bedingung und der 3-Rateversuche-Bedingung (11.0 % vs. 7.2 %), $t(197) = 4.73$, $p < .001$, Cohen's $d = 0.34$. Diese Ergebnisse replizieren einerseits Vorgängerstudien, die zeigen konnten, dass ein einzelner initialer Vortest die Anzahl an Intrusionen im finalen Test reduzieren kann (z. B. Grimaldi & Karpicke, 2012, Kliegl et al., 2024), andererseits erweitern diese Ergebnisse die bisherige Literatur um den Befund, dass auch dieser vorteilhafte Effekt von Vortests noch größer wird, wenn drei Rateversuche statt nur eines einzelnen Rateversuchs abgegeben werden sollen. Zusammen mit den Ergebnissen des korrekten Abrufes legen die Analysen der Intrusionen also nahe, dass wiederholte Rateversuche nicht nur die spätere Erinnerungsleistung für vorgetestetes Material fördern, sondern auch vor Intrusionen schützen können.

Zudem wurde noch eine Analyse der Vortest-Intrusionen – darunter sind fehlerhafte Antworten im finalen Test zu verstehen, die dem selbst generierten Rateversuch für ein gegebenes Cue-Wort entsprechen – durchgeführt. Diese Analyse beinhaltete nur die 1-Rateversuch-Bedingung und die 3-Rateversuche-Bedingung der Experimente 1-4 und zeigte keinen signifikanten Unterschied zwischen den beiden Bedingungen (3.1 % vs. 2.5 %), $F(1,197) = 1.56$, $MSE = 17.32$, $p = .21$, $\eta^2_p = .01$. Numerisch zeigte sich jedoch auch hier ein ähnliches Befundmuster, wie in der oben beschriebenen Gesamtanalyse, mit einer niedrigeren Intrusionsrate bei wiederholten verglichen mit einmaligen Rateversuchen. Die fehlende statistische Signifikanz könnte hier möglicherweise aufgrund eines Bodeneffekts zustande gekommen sein. Dieser Befund mag zudem überraschend wirken, da die Anzahl an potenziellen Vortest-Intrusionen in der 3-Rateversuche-Bedingung dreimal größer ist als in der 1-Rateversuch-Bedingung.

5.5. Diskussion der Experimente 1-4

Die Ergebnisse der vorliegenden Experimente 1-4 liefern erste Hinweise darauf, dass der vorteilhafte Effekt von Vortests auf die spätere Erinnerungsleistung sogar noch größer ausfällt, wenn nicht nur ein einziger, sondern wiederholte Rateversuche während des Lernens des Materials abgegeben werden sollten. Sowohl unter der Verwendung von schwach assoziierten Wortpaaren (Experimente 1 und 3), als auch bei Swahili-Deutsch-Vokabeln (Experiment 2) und Textpassagen (Experiment 4) konnte der klassische Vortest-Effekt, mit einer besseren Erinnerungsleistung für das Lernmaterial in der Vortest-Bedingung verglichen mit der Lern-Bedingung, gezeigt werden. Diese Ergebnisse stellen somit eine Replikation vorheriger Studien dar, die einen Vortest-Effekt sowohl für semantisch verwandtes Lernmaterial wie schwach assoziierte Wortpaare und Textpassagen (z. B. Hays et al., 2013; Kornell, 2014; Richland, 2009) als auch für semantisch nicht verwandte Vokabelpaare (Potts & Shanks, 2014) finden konnten. Vor allem aber nahm die Größe des Vortest-Effekts mit steigender Anzahl an Rateversuchen zu. Dabei zeigte sich nach drei Rateversuchen eine Zunahme des Vortest-Effekts von ca. 35 % (Experiment 4) bis hin zu ca. 100 % (Experimente 1 und 2) verglichen mit der Standard-Vortest-Bedingung mit nur einem einzigem Rateversuch.

In Bezug auf die eingangs gestellte Frage, wie Vortests effektiv in einer angewandten Lernumgebung eingesetzt werden könnten und ob Lehrpersonen den Lernenden nur einmal oder mehrfach die Möglichkeit geben sollten, vor dem eigentlichen Lernen die potenziell korrekte Lösung zu erraten, lässt sich diese Frage auf Grundlage der vorliegenden Ergebnisse mit einer Empfehlung für den Einsatz wiederholter Vortests beantworten. Denn die vorliegenden Ergebnisse deuten darauf hin, dass Vortests insbesondere dann effektiv als Lernstrategie eingesetzt werden können, wenn vor der Präsentation der korrekten Lösung mehrfach falsche Rateversuche generiert werden. Dieser vorteilhafte Effekt wiederholter Vortests hat auch dann noch Bestand, wenn die Lernenden, wie in Experiment 2, (nahezu) keine Chance haben, die korrekte Lösung zu erraten. So könnte beispielsweise ein Spanischlehrer vor der Einführung einer neuen Vokabel die Lernenden zunächst dazu auffordern mehrfach die Bedeutung der neuen Vokabel zu erraten (z. B. „Was heißt „camiseta“ auf Deutsch?“ Korrekte Lösung: „T-Shirt“) oder ein Geografielehrer könnte die Lernenden dazu auffordern, mehrfach zu versuchen, die Hauptstadt eines Landes zu erraten (z. B. „Was ist die Hauptstadt von Mosambik?“ Korrekte Lösung: „Maputo“), bevor die korrekte Lösung präsentiert wird.

Der Befund, dass wiederholte Vortests, verglichen mit nur einem einzigen Vortest, sogar noch zu einer Vergrößerung des Erinnerungsvorteils für das vorgetestete Material führten, was sich in dem größeren Vortest-Effekt widerspiegelt, ist konsistent mit der Elaborations-Hypothese des Vortest-Effekts. Denn gemäß diesem Erklärungsansatz soll das Generieren eigener Rateversuche im Vortest zu einer umfangreichen Aktivierung von Gedächtnisrepräsentationen führen, die mit dem Cue-Item verwandt sind. Diese verwandten Informationen sollen wiederum mit dem Cue-Target-Wortpaar verknüpft werden und könnten in einem finalen Test als semantische Mediatoren fungieren und dadurch den Abruf des korrekten Target-Wortes erleichtern. Basierend auf den Annahmen der Elaborations-Hypothese ergibt sich die Vermutung, dass wiederholtes Vortesten folglich zu einer noch umfangreicher elaborierten Gedächtnisspur führen sollte, als dass das bei nur einem Rateversuch der Fall sein sollte. Denn einerseits sollte dadurch noch mehr verwandte Information mit dem Cue-Target-Wortpaar verknüpft werden und andererseits sollte die Anzahl an fehlerhaften Rateversuchen, die im finalen Test als mögliche semantische

Mediatoren fungieren könnten, bei wiederholten Rateversuchen doppelt bzw. dreimal so groß sein sollte wie bei nur einem Rateversuch.

So wäre es beispielsweise denkbar, dass eine Versuchsperson in der 1-Rateversuch-Bedingung bei der Präsentation des Cue-Wortes *Garage* fälschlicherweise das Wort *Auto* rät, das bei der Präsentation der korrekten Antwort *Haus* mit dieser verknüpft wird. In einem finalen Test könnte bei der Präsentation des Cue-Wortes zunächst wieder das Wort *Auto* abgerufen werden, das den Zugang zum Target-Wort *Haus* erleichtert, da sich die Versuchsperson möglicherweise den Zusammenhang „wenn ich mit dem Auto in der Garage bin, bin ich in meinem Haus“ gemerkt hat. Generiert die Versuchsperson nun aber in der 3-Rateversuche-Bedingung neben dem Rateversuch *Auto* auch noch die Rateversuche *Tor* und *Einfahrt*, stellt die Versuchsperson möglicherweise die zusätzlichen Verknüpfungen „das Garagentor führt zu meinem Haus“ und „wenn ich die Einfahrt zur Garage hochfahre, ist da ein Haus“ her, wodurch sich die Chance erhöht, durch einen semantischen Mediator an das korrekte Target-Wort *Haus* erinnert zu werden. Dies sollte zu einer insgesamt besseren Erinnerungsleistung in der 3-Rateversuche-Bedingung als in der 1-Rateversuch-Bedingung führen. Insgesamt sollte also gemäß der Elaborations-Hypothese der Vortest-Effekt durch das wiederholte Generieren eines Rateversuchs größer ausfallen, als wenn nur ein einziger Rateversuch abgegeben wird, was mit den Ergebnissen der Experimente 1-4 gezeigt werden konnte. Für die Verwendung von Swahili-Deutsch-Vokabeln als Lernmaterial allerdings scheint die Erklärung mittels der Elaborations-Hypothese schwierig, da hier weder zwischen dem Cue-Wort und dem Target-Wort noch zwischen dem Cue-Wort und den generierten Rateversuchen eine präexistierende semantische Verknüpfung bestehen sollte. Durch diese ausbleibende Verknüpfung sollten die Rateversuche in einem späteren finalen Test nur bedingt den Zugang zu dem korrekten Target-Wort erleichtern, weshalb die Befunde aus Experiment 2 nur schwer mit den Annahmen der Elaborations-Hypothese zu vereinbaren sind.

Neben der Elaborations-Hypothese passen die vorliegenden Ergebnisse aber auch zu den Annahmen der zweiten prominenten Theorie des Vortest-Effekts, der Aufmerksamkeits-Hypothese. Denn diese geht davon aus, dass das Generieren eines Rateversuchs im Vortest zu einer aufmerksameren Enkodierung des nachfolgenden Feedbacks führt (z. B. Potts & Shanks, 2014), da die Versuchspersonen durch den Vortest auf ein Informationsdefizit

hingewiesen werden und dadurch bestrebt sind, diese Lücke durch die korrekte Information zu füllen. Diese Neugierde und die damit verbundene Verwendung von Ressourcen für die Verarbeitung der Target-Items könnte umso größer ausgefallen sein, je mehr Rateversuche die Versuchspersonen selbst abgegeben haben, was sich in der beobachteten Vergrößerung des Vortest-Effekts widerspiegeln könnte.

Zudem lassen sich die gezeigten Befunde auch noch mit einem weiteren Erklärungsansatz vereinbaren. So geht die episodische Kontexthypothese des Vortest-Effekts (Metcalf & Huelser, 2020) davon aus, dass bei der Enkodierung von Information auch Merkmale des räumlich-zeitlichen Kontexts, die während des Enkodierens vorliegen, unbewusst mit dieser Information zusammen abgespeichert werden (Estes, 1955; Mensik & Raaijmakers, 1988). Während Items in der Lern-Bedingung also nur mit dem Lern-Kontext verknüpft sein sollten, könnten Items in der Vortest-Bedingung hingegen mit einem umfangreicheren Kontext assoziiert sein, der aus Merkmalen aus dem Vortest-Kontext und auch aus dem anschließenden Lern-Kontext besteht. Items in der Bedingung mit wiederholten Rateversuchen könnten nun folglich nicht nur mit Merkmalen aus einem Vortest-Kontext, sondern mit kontextuellen Merkmalen aus mehreren neuen, zeitlichen Vortest-Kontexten verknüpft sein, was zu einer insgesamt noch größeren Vielfalt an Kontextmerkmalen führen sollte. Da diese Kontextmerkmale, ähnlich wie semantische Mediatoren, in einem finalen Test reaktiviert werden und damit den Zugang zu den Target-Items erleichtern sollten, könnte die größere Verfügbarkeit an unterschiedlichen hilfreichen Kontextmerkmalen bei wiederholten Vortests zu einem noch größeren Erinnerungsvorteil geführt haben als bei einmaligen Vortests.

Die Search-Set-Hypothese (Grimaldi & Karpicke, 2012) nimmt an, dass bei der Bearbeitung eines Vortests ein Search-Set aktiviert wird, das unterschiedliche Items, einschließlich der korrekten Antwort, umfasst. Durch die Voraktivierung des Target-Items in diesem Search-Set sollte diese Zielinformation bei der Präsentation des korrektiven Feedbacks effektiver enkodiert werden. Jeder Vortest bzw. Rateversuch während des Vortests sollte ein weiteres, leicht unterschiedliches Search-Set aktivieren, was die Wahrscheinlichkeit erhöht, dass das korrekte Target-Item voraktiviert und nachfolgend besser abgespeichert wird. Da diese tiefere Enkodierung der Zielinformation im finalen Test den Abruf des Target-Items erleichtern und zu einer besseren Erinnerungsleistung beitragen

sollte, wäre gemäß der Search-Set-Hypothese zu erwarten, dass sich bei wiederholten Rateversuchen im Vergleich zu nur einem Rateversuch während des Vortests ein noch größerer Vortest-Effekt zeigt. Diese Erwartung stimmt mit den Ergebnissen der Experimente 1-4 überein.

Während die vorliegenden Ergebnisse also mit den Annahmen unterschiedlicher theoretischer Erklärungsansätze in Einklang stehen, lassen sich die Befunde nicht mit dem Cue-Overload-Prinzip erklären. Gemäß diesem Ansatz sollte ein Rateversuch mit dem Cue-Item verknüpft werden und somit in einem späteren finalen Test eine Konkurrenz zum korrekten Target-Item darstellen und den Abruf dieses Target-Items erschweren. Mehrere Rateversuche würden dementsprechend noch mehr Verknüpfungen mit dem Cue-Item und damit noch mehr Konkurrenz beim Abruf des Target-Items bedeuten, was im finalen Test bei mehreren Rateversuchen verglichen mit der 1-Rateversuch-Bedingung einerseits zu einer schlechteren Erinnerungsleistung für die Target-Items und andererseits zu mehr Intrusionen führen sollte. Die vorliegenden Ergebnisse aller Experimente zeigen jedoch, dass wiederholtes Raten zu einer besseren Erinnerungsleistung und weniger Intrusionen führt als das einmalige Raten bzw. das bloße Lernen des Materials. Dies deutete darauf hin, dass das Generieren mehrerer Rateversuche den Zugang zu den Target-Items zu erleichtern scheint, statt diesen - wie vom Cue-Overload-Prinzip vorhergesagt - zu erschweren. Die Befunde implizieren allerdings nicht, dass es durch Vortests nicht zu einem Cue-Overload kommen kann, sondern diese zeigen lediglich, dass dies bei bis zu drei generierten Rateversuchen noch nicht der Fall zu sein scheint. Was allerdings bei noch steigender Anzahl an Rateversuchen (beispielsweise fünf oder zehn Rateversuchen) passiert, bleibt noch offen. Hier wäre es jedoch denkbar, dass es ab einer gewissen Anzahl an Vortests durch eine Überladung des Cues und einen erschwerten Abruf des Target-Items zu einer stärkeren Beeinträchtigung der Erinnerungsleistung und einem reduzierten oder ausbleibenden Vortest-Effekt kommt.

Ebenso wie mit dem Cue-Overload-Prinzip lassen sich die gefundenen Ergebnisse auch nur schwer mit der Error-Correction-Hypothese (Carrier & Pashler, 1992) in Einklang bringen. Gemäß diesem theoretischen Erklärungsansatz sollte sich die Anzahl der abgegebenen Rateversuche nicht auf die Erinnerungsleistung und auf die Größe des Vortest-Effekts auswirken. Denn nach der Error-Correction-Hypothese sollte die Präsentation der

korrekten Lösung unabhängig von der Anzahl der falschen Rateversuche ein Fehlersignal auslösen. Die dadurch aktivierten Fehlerkorrekturmechanismen führen dann zu einer Stärkung der Abrufroute zwischen dem Cue-Item und dem Target-Item. Auf Grundlage dieser Theorie wäre zwar insgesamt eine bessere Erinnerungsleistung in der Vortest-Bedingung als in der Lern-Bedingung zu erwarten, jedoch sollten sich keine Unterschiede zwischen den Vortest-Bedingungen mit einem oder mehreren Rateversuchen zeigen. In allen vier Experimenten konnte hingegen eine Zunahme des Vortest-Effekts mit steigender Anzahl generierter Rateversuche gefunden werden.

Falls in den vorliegenden Experimenten tatsächlich Elaborationsprozesse für die Entstehung des Vortest-Effekts bei der vorgetesteten Information verantwortlich waren und sich diese Elaboration auch auf ungetestete Information übertragen haben sollte, dann sollte der Abruf der ungetesteten Information nach drei Rateversuchen besser gelingen als nach einem einzigen Rateversuch. Tatsächlich zeigen die Ergebnisse von Experiment 4 dieses Muster, was zunächst die Elaborations-Hypothese zu stützen scheint. Gemäß dieser Argumentation sollte die Erinnerungsleistung an die ungetestete Information nach einem einzigen Rateversuch, vor allem aber nach drei Rateversuchen, auch höher sein als in der Lern-Bedingung, was in den Ergebnissen des vorliegenden Experiments 4 nicht der Fall war. Vielmehr scheinen weder ein Rateversuch noch drei Rateversuche den Abruf ungetesteter Information fördern zu können, was nahelegt, dass sich die durch Vortests ausgelösten Erinnerungsvorteile nicht direkt auf gelernte, aber ungetestete Information in diesen Bedingungen transferieren lassen. Die Tatsache, dass die Erinnerungsleistung für ungetestete Informationen in der 1-Rateversuch-Bedingung sogar signifikant schlechter ausfiel als in der Lern-Bedingung, steht im Widerspruch zu bisherigen Studien, die keine negativen Effekte von Vortests auf den Abruf ungetesteter Information finden konnten (z. B. James & Storm, 2019). Ein möglicher Grund für den negativen Effekt eines einzigen Rateversuchs auf den Abruf von ungetesteter Information in Experiment 4 könnte gewesen sein, dass die Versuchspersonen für Sätze mit Vortests so viel Aufmerksamkeit verwendeten, dass sie die unmittelbar danach präsentierte Information in den Sätzen ohne Vortest, in Erwartung des nächsten Vortest-Satzes, weitgehend ignorierten und dementsprechend zu einem späteren Zeitpunkt schlechter abrufen konnten.

Ein weiterer wichtiger Befund der vorliegenden Experimente war die Tatsache, dass wiederholte Rateversuche während der Lern-Phase zu einer Vergrößerung des Vortest-Effekts führten, unabhängig davon, ob die Dauer eines Lern-Trials mit steigender Anzahl an Rateversuchen zunahm (Experimente 1, 2, 4) oder ob diese, wie in Experiment 3, für die unterschiedliche Anzahl an Rateversuchen konstant gehalten wurde. Diese Befunde legen nahe, dass es tatsächlich die Anzahl der Rateversuche und nicht die längere Auseinandersetzung mit dem Lernmaterial zu sein scheint, die für den Anstieg des Vortest-Effekts verantwortlich ist. Diese Ergebnisse passen gut zu einer Vorgängerstudie von Kornell und Vaughn (2016), die in ihren Experimenten zeigen konnten, dass die bloße Verlängerung der Zeitspanne von 5 s auf bis zu 30 s, die den Versuchspersonen in Vortest-Trials zur Generierung eines Rateversuchs zur Verfügung stand, nicht zu einer Vergrößerung des Vortest-Effekts führte.

Zusammenfassend lässt sich also festhalten, dass durch die Betrachtung des Einflusses der Testanzahl auf die Größe des Vortest-Effekts in den Experimenten 1-4 ein wichtiger Beitrag zur Schließung dieser empirischen Lücke geleistet werden konnte. Zudem lassen sich ausgehend von den gezeigten Befunden Impulse für den effizienten und gewinnbringenden Einsatz von Vortests in Lernsituationen ableiten. Darüber hinaus liefern die Ergebnisse dieser Experimentserie Hinweise auf zugrundeliegende Mechanismen des Vortest-Effekts. Dabei scheinen durch die Bearbeitung von Vortests vor allem semantische Elaborationsprozesse sowie Aufmerksamkeitsprozesse ausgelöst zu werden, die zu einer besseren Erinnerungsleistung in einem späteren Test beitragen. Ausgehend von diesen Befunden soll im Anschluss gezielt und separat betrachtet werden, welche Rolle diese beiden Prozesse bei der Entstehung des Vortest-Effektes spielen. Zu diesem Zweck wurden die beiden nachfolgenden Experimentserien durchgeführt. Dabei widmeten sich die Experimente 5 bis 7 der Untersuchung der Elaborationsprozesse, wobei konkrete Vorhersagen der Elaborations-Hypothese überprüft werden sollten. Die Experimente 8 und 9 hingegen dienten der Überprüfung der Annahmen der Aufmerksamkeits-Hypothese.

6. Die Rolle von semantischen Mediatoren für Test-Effekte

6.1. Bisherige Befunde

Die Frage, ob und wenn ja, welche Art der Information während der Durchführung einer Abrufübung aktiviert wird und durch welche Prozesse sich diese potenzielle Aktivierung erklären lässt, beschäftigt Forscher seit vielen Jahren. Einen Forschungszweig stellt dabei die Untersuchung von Elaborationsprozessen dar, für deren experimentelle Überprüfung die Verwendung semantischer Mediatoren, d. h. Informationen, die mit dem Lernmaterial verwandt sind, aber selbst nicht gelernt werden, eine wichtige Rolle spielt. So gibt es mehrere Studien, in denen der Einfluss semantischer Mediatoren auf den klassischen Test-Effekt betrachtet wurde. Eine Studie stammt dabei von Pyc und Rawson (2010), in deren Experiment Versuchspersonen Swahili-Englisch Vokabeln lernen sollten (z. B. Samadi - Dünger), die entweder im Anschluss erneut gelernt werden konnten oder für die im Anschluss ein Cued-Recall-Test (Samadi - ?) bearbeitet werden sollte. Anschließend fand in beiden Bedingungen ein finaler Test statt. Dabei wurden die Versuchspersonen aufgefordert, während des erneuten Lernens bzw. während des ersten Abrufs der Information einen semantischen Mediator zu generieren, der eine Verbindung zwischen dem Cue-Wort und dem Target-Wort herstellen und den Versuchspersonen in einem späteren finalen Test den Zugang vom Cue-Wort zum Target-Wort erleichtern sollte (z. B. Samadi -> „Samen“ -> Dünger). Die Ergebnisse des finalen Cued-Recall-Tests ergaben dabei nicht nur eine bessere Erinnerungsleistung in der Test-Bedingung verglichen mit der erneuten Lern-Bedingung, also den klassischen Test-Effekt, sondern auch eine bessere Erinnerungsleistung für den selbstgenerierten Mediator in dieser Bedingung. Die Bearbeitung einer Abrufübung führte also nicht nur zu einer Stärkung des Gedächtnisses für die Lerninhalte, sondern förderte auch die Nutzung von und die Erinnerung an die semantischen Mediatoren. Daraus lässt sich die Frage ableiten, ob die Verfügbarkeit der semantischen Mediatoren für die Entstehung des Test-Effekts (mit)verantwortlich ist.

Ausgehend von der Vorgängerstudie von Pyc und Rawson (2010) widmete sich Carpenter (2011) in zwei Experimenten dieser Fragestellung. Hier sollten die Versuchspersonen schwach assoziierte Cue-Target-Paare (z. B. Jacke – Shirt) lernen, die nach der Lernphase entweder erneut gelernt wurden oder für die ein Test bearbeitet werden sollte. In Experiment 1 fand abschließend ein Rekognitionstest statt, bei dem sowohl die Cue-Wörter und die Target-Wörter als auch neue unbekannte Items präsentiert wurden. Die unbekannten Wörter bestanden dabei entweder aus unverwandten Wörtern (z. B. Leiter) oder aus semantisch mit dem Cue-Wort verwandten Wörtern (z. B. Mantel), den semantischen Mediatoren. Die Ergebnisse zeigten dabei einerseits eine bessere Erinnerungsleistung für die Cue- und Target-Wörter in der Test-Bedingung als in der Lern-Bedingung und andererseits wurden in der Test-Bedingung signifikant mehr semantische Mediatoren fälschlicherweise als „bekannt“ eingeschätzt. In Experiment 2 fand der finale Test im Cued-Recall-Format statt, wobei die Versuchspersonen versuchen sollten, gelernte Target-Wörter abzurufen, während entweder das dazugehörige Cue-Wort oder aber ein semantischer Mediator als Cue-Wort präsentiert wurde. Bei der Betrachtung der Ergebnisse zeigte sich für die Verwendung der originalen Cue-Items der klassische Test-Effekt. Überraschenderweise führte aber auch die Präsentation eines semantischen Mediators als Cue-Item in der Test-Bedingung zu einer besseren Erinnerung als in der Lern-Bedingung. Diese Ergebnisse wurden von der Autorin als Hinweis dahingehend interpretiert, dass es während der Bearbeitung eines Tests auch ohne die gezielte Aufforderung, einen Mediator zu genieren, zu einer Aktivierung semantisch verwandter Information kommt, die mit dem Cue-Target-Paar verknüpft wird und so in einem finalen Test den Zugang zum Target-Item erleichtert.

Der Versuch einer Replikation der Befunde von Carpenter (2011) stammt aus einer neueren Studie von Di Marco und Marmurek (2022), die ebenfalls den Einfluss semantischer Mediatoren auf den Test-Effekt betrachteten. Auch in deren Experimenten sollten Versuchspersonen Wortpaare lernen, die anschließend entweder erneut gelernt oder getestet wurden, bevor in einem finalen Cued-Recall-Test semantische Mediatoren präsentiert wurden und die gelernten Target-Wörter abgerufen werden sollten. Die Ergebnisse zeigten dabei auch hier eine bessere Erinnerungsleistung für Target-Items der Test-Bedingung verglichen mit Target-Items der Lern-Bedingung, womit sich die Befunde von Carpenter (2011) replizieren ließen.

Insgesamt lässt sich also festhalten, dass die bisherigen Studien zur Untersuchung des Einflusses semantischer Mediatoren auf die vorteilhaften Effekte von Abrufübungen auf die spätere Erinnerungsleistung ausschließlich unter Verwendung des klassischen Test-Effekt-Paradigmas durchgeführt wurden. Hierbei zeigte sich auch dann ein Test-Effekt, wenn statt des ursprünglich gelernten Cue-Wortes im finalen Test ein semantischer Mediator verwendet wurde, was nahelegt, dass Elaborationsprozesse für die Entstehung des klassischen Test-Effekts eine Rolle spielen könnten. Unklar ist jedoch, inwiefern sich diese Befunde auch auf die Verwendung von Vortests als Art der Abrufübung übertragen lassen und inwiefern auch hier Elaborationsprozesse eine entscheidende Rolle spielen könnten. Zwar könnte davon ausgegangen werden, dass sich die Befunde von Carpenter (2011) aufgrund methodischer Ähnlichkeiten einfach vom klassischen Test-Effekt auf den Vortest-Effekt übertragen lassen, allerdings würde diese Vermutung einen zentralen Unterschied der beiden Paradigmen vernachlässigen. So liegt der Anteil an Fehlern während der Abrufübung beim klassischen Test-Effekt zwischen 50 % und 0 %, wohingegen diese Fehlerquote beim Vortest-Effekt meist nahezu 100 % beträgt. Dieser Unterschied in der Fehlerrate könnte beeinflussen, wodurch der Erinnerungsvorteil in den beiden Paradigmen zustande kommt. Es wäre also denkbar, dass diesen beiden Arten der Abrufübung folglich auch unterschiedliche kognitive Prozesse zugrunde liegen. Um diese Prozesse näher zu untersuchen, soll in den folgenden Experimenten 5 bis 7 der Einfluss semantischer Mediatoren auf die Größe des Vortest-Effekts betrachtet werden. Aufgrund mangelnder Vorbefunde soll der folgende Abschnitt dazu dienen, basierend auf unterschiedlichen Erklärungsansätzen des Vortest-Effekts, Vermutungen für mögliche Resultate dieser Manipulation auf die Größe des Vortest-Effekts abzuleiten.

6.2. Theoretische Erwartungen

In der Studie von Carpenter (2011) wurden die von der Autorin gefundenen Ergebnisse mit der Elaborations-Hypothese des Test-Effekts erklärt, wobei es durch den Abrufprozess im Test zu einer Integration verwandter Information in das bestehende, gelernte Cue-Target-Paar kommt, die in einem späteren Test den Abruf der Target-Information erleichtert. Auch für den Vortest-Effekt wird eine Version der Elaborations-Hypothese als Erklärungsansatz verwendet, aus der sich bezüglich der Verwendung

semantischer Mediatoren konkrete Vorhersagen ergeben. So geht die Elaborations-Hypothese (Grimaldi & Karpicke, 2012) davon aus, dass es bei der Bearbeitung eines Vortests durch das Generieren eines Rateversuchs zu einer Aktivierung eines semantischen Netzwerks kommt. Dieses Netzwerk beinhaltet dabei Informationen, die mit dem präsentierten Cue-Wort verwandt sind. Daraus lässt sich die Erwartung ableiten, dass durch diese Aktivierung in der Vortest-Bedingung die Wahrscheinlichkeit steigt, dass ein anderes, mit dem Cue-Wort verwandtes Wort, wie beispielsweise ein semantischer Mediator, in einem späteren Rekognitionstest als bekannt oder vertraut eingeschätzt wird. Diese Erwartung ergibt sich, da dieser semantische Mediator wiederum eng mit der aktivierten Information verknüpft ist oder dieser möglicherweise sogar entspricht. In einer reinen Lern-Bedingung hingegen findet diese Voraktivierung nicht statt, weshalb semantische Mediatoren hier nicht häufiger als bekannt eingeschätzt werden sollten. So wäre es beispielsweise denkbar, dass die Präsentation des Cue-Wortes *Ball* zu einer Aktivierung der verwandten Informationen *Sport*, *Pokal*, *Mannschaft* führt. Diese Information wiederum könnte mit dem später präsentierten Mediator-Item *Wettbewerb* verknüpft sein. Aufgrund dieser Verknüpfung sollte die Wahrscheinlichkeit steigen, dass der Mediator *Wettbewerb* in einem späteren Rekognitionstest als bekannt eingeschätzt wird. Würde alternativ das Item *Sport* in einem späteren Rekognitionstest als Mediator präsentiert werden, sollte auch dieses Item wahrscheinlicher als bekannt bewertet werden, da es bereits während des Vortests aktiviert wurde.

Zusätzlich geht die Elaborations-Hypothese davon aus, dass es nicht nur zu einer Aktivierung verwandter Information kommt, sondern diese auch mit in das Cue-Target-Paar integriert wird, sobald das korrektive Feedback in der Lernphase präsentiert wird. Dadurch sollte sowohl die Verknüpfung zwischen dem Cue-Wort (*Ball*) und der verwandten Information (*Sport*, *Pokal*, *Mannschaft*) gestärkt werden, als auch eine Verknüpfung zwischen der Target-Information (*Stadion*) und dem semantischen Mediator entstehen. In einem finalen Test könnte dann die erneute Präsentation des ursprünglichen Cue-Wortes (*Ball*) zunächst zu einer Reaktivierung der semantisch verwandten Information führen, die wiederum den Zugriff auf die gesuchte Target-Information (*Stadion*) erleichtert. Wird im finalen Cued-Recall-Test hingegen ein semantischer Mediator (z. B. *Sport*) als Cue-Item verwendet, sollte dieser in der Vortest-Bedingung mit höherer Wahrscheinlichkeit zu einem

korrekten Abruf des Target-Items (*Stadion*) führen, da dieser während des Vortests bereits aktiviert und mit dem Target-Item verknüpft wurde. In der Lern-Bedingung sollte es keine derartige Verknüpfung geben, weshalb hier der Abruf der Target-Information schwerer fallen sollte. Insgesamt sollte sich basierend auf der Elaborations-Hypothese sowohl für die Verwendung des ursprünglichen Cue-Wortes als auch für die Präsentation eines semantischen Mediators eine bessere Erinnerung an die korrekte Target-Information in der Vortest-Bedingung zeigen, was die Rolle von Elaborationsprozessen für die Entstehung des Vortest-Effekts betonen würde.

Eine andere Erwartung bezüglich des Einflusses semantischer Mediatoren auf den Vortest-Effekt liefert die Error-Correction-Hypothese (Carrier & Pashler, 1992). Dieser Erklärungsansatz geht davon aus, dass das Generieren eines fehlerhaften Rateversuchs und die unmittelbar darauffolgende Präsentation der korrekten Lösung in der Vortest-Bedingung zu einem Fehlersignal führt. Dieses Fehlersignal führt einerseits zu einer Stärkung der Verknüpfung zwischen dem Cue-Item und dem Target-Item, wodurch die korrekte Abrufroute gefestigt wird (Knight et al., 2012). Andererseits soll es durch die Aktivierung von Fehlerkorrekturmechanismen zu einer Abschwächung der Verknüpfung des Cue-Items mit dem inkorrekten Rateversuch sowie zu einer Hemmung zusätzlicher, für den Abruf des Target-Items hinderlicher Information, wie beispielsweise semantisch verwandter Information, kommen. Im Gegensatz zur Elaborations-Hypothese, die von einer Integration semantisch verwandter Information, einschließlich des fehlerhaften Rateversuchs, in die Cue-Target-Verknüpfung ausgeht, wird diese Information gemäß der Error-Correction-Hypothese unterdrückt und gehemmt.

Daraus ergibt sich die Erwartung, dass die Vertrautheit der Cue- und Target-Items in der Vortest-Bedingung gesteigert und diese in einem Rekognitionstest mit größerer Wahrscheinlichkeit korrekt wiedererkannt werden sollten als in der Lern-Bedingung. Unbekannte neue Informationen, wie semantische Mediatoren, sollten in der Vortest-Bedingung allerdings der Hemmung durch Fehlerkorrekturmechanismen unterliegen und daher in einem Rekognitionstest weder häufiger noch seltener als bekannt eingeschätzt werden als in der Lern-Bedingung. In einem finalen Cued-Recall-Test sollte die Präsentation des ursprünglichen Cue-Items durch die Stärkung der korrekten Abrufroute in der Vortest-Bedingung zu einer besseren Abrufleistung für die Target-Items führen. Wird jedoch ein

semantischer Mediator als Cue-Item präsentiert, sollte dieser aufgrund der Hemmung aller zusätzlicher, nicht zielführender Informationen in der Vortest-Bedingung nicht zu einem Erinnerungsvorteil für die Target-Items führen.

Aus anderen prominenten Erklärungsansätzen lassen sich bezüglich des Einflusses semantischer Mediatoren auf den Vortest-Effekt keine konkreten Erwartungen ableiten. So sollte weder eine gesteigerte Aufmerksamkeit bei der Enkodierung der Target-Items (Aufmerksamkeits-Hypothese) noch die Integration von räumlich-zeitlichen Kontextmerkmalen in das Lernmaterial (Kontext-Hypothese) oder die bessere Enkodierung der Target-Informationen durch die Voraktivierung der Items (Search-Set-Hypothese) einen Einfluss auf die Rolle semantischer Mediatoren haben. Der Beitrag dieser Erklärungsansätze für den Vortest-Effekt lässt sich anhand der nachfolgenden Experimente daher nicht oder nur bedingt überprüfen. Folglich werden die nachfolgenden Experimente 5 bis 7 insbesondere zur Untersuchung der unterschiedlichen Erwartungen der Error-Correction-Hypothese und der Elaborations-Hypothese bezüglich des Einflusses semantischer Mediatoren auf den Vortest-Effekt durchgeführt.

6.3. Fragestellung der Experimente 5-7

In den bisherigen Studien zum Vortest-Effekt erfolgte das Lernen stets intentional, wobei die Versuchspersonen eine konkrete Lerninstruktion erhielten und darüber informiert wurden, dass ein späterer Abruf der Inhalte stattfindet. Bislang ist daher noch unklar, ob Lernende noch immer von Vortests profitieren, wenn der Lernprozess unbewusst geschieht, weshalb in den Experimenten 5-7 inzidentelles Lernen verwendet wurde, um diese Frage zu beantworten. Außerdem wurden in den nachfolgenden Experimenten neben den ursprünglich gelernten Cue-Items im finalen Test auch neue unbekannte, aber mit dem Lernmaterial verwandte Cue-Items eingesetzt. Damit sollte untersucht werden, ob Vortests auch dann zu einer besseren Erinnerung an die Target-Items beitragen können, wenn semantische Mediatoren als Abruf-Cues genutzt werden. Diese Fragestellung ist von großer Relevanz, da Lernende beim Abrufen von Informationen häufig mit Transfer-Fragen oder neuen Abruf-Cues konfrontiert werden. Ein Ziel der Experimente 5 bis 7 bestand folglich darin, die Literatur zum Vortest-Paradigma zu erweitern und neue Hinweise bezüglich der Generalisierbarkeit der positiven Effekte von Vortests auf die spätere Erinnerungsleistung

zu liefern. Aus den Ergebnissen lassen sich insbesondere wichtige Erkenntnisse über die Voraussetzungen des effektiven Einsatzes des Vortest-Paradigmas im angewandten pädagogischen Kontext ableiten.

Studien zum klassischen Test-Effekt liefern bereits einige Hinweise darauf, dass Elaborationsprozesse eine Rolle für die Entstehung von Abruf-Effekten zu spielen scheinen. Ebenso legen auch die Ergebnisse der Experimente 1 bis 4 nahe, dass die gefundene Vergrößerung des Vortest-Effekts unter anderem auf Elaborationsprozesse zurückzuführen gewesen sein könnte. Es gibt bisher jedoch noch kaum Studien, die gezielt untersuchten, inwiefern semantische Elaborationsprozesse eine zentrale Rolle für den Vortest-Effekt spielen, wie dies die Elaborations-Hypothese für den Vortest-Effekt (Carpenter, 2009) postuliert. Daher war es ein weiteres Ziel der Experimente 5-7 des nachfolgenden Kapitels, genau diese Fragestellung näher zu betrachten und die Elaborations-Hypothese bezüglich ihres Erklärungsgehalts für den Vortest-Effekt zu überprüfen. In drei Experimenten, die auf der Vorgängerstudie von Carpenter (2011) aufbauten, aber statt des klassischen Test-Effekt-Paradigmas ein Vortest-Paradigma verwendeten, sollten diese semantischen Elaborationsprozesse getestet werden. Dabei sollten die Versuchspersonen in einer anfänglichen Lernphase schwach assoziierte Wortpaare (z. B. Mutter – Kind) lernen, von denen die Hälfte der Items vorgetestet und die andere Hälfte der Items direkt gelernt wurden. In Anlehnung an Experiment 1 aus der Studie von Carpenter (2011) wurde in Experiment 5 abschließend ein Rekognitionstest verwendet, um die Erinnerungsleistung zu bestimmen. Dabei wurden sowohl die alten Cue- und Target-Items als auch neue mit dem restlichen Lernmaterial unverwandte Items und semantisch mit dem Cue verwandte Mediator-Items verwendet (z. B. Vater). In den Experimenten 6 und 7 fand wie in Carpenter's Experiment 2 der finale Test im Cued-Recall-Format statt, wobei die Versuchspersonen einen Abruf-Cue präsentiert bekamen und ein Target-Item abrufen sollten. Dabei wurde der Hälfte der Versuchspersonen in beiden Experimenten das originale Cue-Item als Abruf-Cue präsentiert (z. B. Mutter - ?) und der anderen Hälfte ein semantischer Mediator (z. B. Vater - ?). Die Experimente 6 und 7 waren dabei identisch mit der einzigen Ausnahme, dass die Dauer des Retentionsintervalls zwischen der Lern-Phase und dem finalen Test von 20 min (Experiment 6) auf 4h (Experiment 7) verlängert wurde.

Wenn sich die von Carpenter (2011) für den klassischen Test-Effekt gefundenen Ergebnisse auch auf das Paradigma des Vortest-Effekts generalisieren lassen, sollte sich in Experiment 5 eine größere Hit Rate für Target-Items in der Vortest-Bedingung verglichen mit der Lern-Bedingung zeigen und im finalen Cued-Recall-Test der Experimente 6 und 7 ein höherer Anteil an korrekt erinnerten Target-Wörtern in der Vortest-Bedingung als in der Lern-Bedingung zu finden sein. Es wäre also in allen drei Experimenten ein klassischer Vortest-Effekt zu erwarten. Vor allem aber sollte eine Generalisierbarkeit der Befunde von Carpenter auch zu einer höheren False Alarms Rate für semantische Mediatoren in der Vortest-Bedingung verglichen mit der Lern-Bedingung (Experiment 5) führen. Zudem wäre auch eine bessere Erinnerungsleistung für die Target-Items der Vortest-Bedingung verglichen mit den Target-Items der Lern-Bedingung zu erwarten, wenn ein semantischer Mediator als Abruf-Cue präsentiert wird (Experimente 6 und 7).

Ein derartiges Ergebnismuster würde die Elaborations-Hypothese als Erklärungsansatz für den Vortest-Effekt stützen, denn gemäß dieser Hypothese sollte die Aktivierung zusätzlicher, semantisch mit dem Cue verwandter Information während des Generierens eines Rateversuchs im Vortest dazu führen, dass diese zusätzliche Information (z. B. semantische Mediatoren) von den Versuchspersonen als vertrauter eingeschätzt wird. In einem finalen Rekognitionstest sollten diese Mediatoren dann mit größerer Wahrscheinlichkeit fälschlicherweise als bekannte Items eingeordnet werden, was sich in einer höheren False Alarms Rate zeigen sollte (Experiment 5). Des Weiteren könnte die Aktivierung der zusätzlichen Information durch den Vortest zu einer Verknüpfung dieser semantisch verwandten Information mit dem bestehenden Cue-Target-Paar führen, was in einem späteren Cued-Recall-Test bei der Präsentation des semantischen Mediators als Abruf-Cue den Zugang zum Target-Item erleichtern und somit zu einer besseren Erinnerungsleistung für die Target-Items führen sollte, die in der Vortest-Bedingung verglichen mit der Lern-Bedingung gelernt wurden (Experiment 6 und 7).

Auf Grundlage der Error-Correction-Hypothese hingegen wäre zu erwarten, dass sich weder die False Alarms Rate für Mediator-Items noch die Erinnerungsleistung für Target-Items bei der Präsentation von Mediatoren als Cue-Items zwischen der Vortest- und der Lern-Bedingung unterscheidet. Vielmehr wäre es denkbar, dass diese beiden Indikatoren in der Vortest-Bedingung sogar geringer ausfallen, wenn Fehlerkorrekturmechanismen nicht

nur zu einer Unterdrückung der Verknüpfung zwischen dem Cue-Wort und dem Rateversuch, sondern auch zu einer Hemmung der Verknüpfung des Cue-Wortes mit anderer verwandter Information, wie beispielsweise Mediatoren, führen. Die nachfolgenden Experimente sollten folglich erste Erkenntnisse darüber liefern, inwiefern Elaborationsprozesse oder Fehlerkorrekturmechanismen für die Entstehung des Vortest-Effekts eine wichtige Rolle spielen. Dementsprechend stellen die nachfolgenden Experimente einen Test konkreter Annahmen der Elaborations-Hypothese und der Error-Correction-Hypothese dar.

7. Die Rolle von Mediatoren für den Vortest-Effekt

7.1. Experiment 5

Method

Ethische Berücksichtigungen. Alle nachfolgend berichteten Experimente 5-7 wurden in Übereinstimmung mit den Vorgaben der Deklaration von Helsinki des Weltärztebundes durchgeführt.

Stichprobe. Zur Bestimmung der notwendigen Stichprobengröße für die Experimente 5-7 wurde eine Poweranalyse mittels G*Power (Version 3.1.9.2; Faul et al., 2009) durchgeführt. Als Parameter wurden hierbei ein α von .05, ein β von .20 und eine Effektgröße von Hedges' g von 0.44 basierend auf der metaanalytischen Schätzung der Größe des Vortest-Effekts (Boustani & Shanks, 2022) verwendet, wobei sich eine notwendige Stichprobengröße von $N = 43$ pro Zwischensubjektfaktor ergab. Die tatsächliche Stichprobe war mit 54 Versuchspersonen ($M = 24.1$ Jahre, Altersspanne = 18-35 Jahre, 45 weiblich) etwas größer. Alle Versuchspersonen sprachen Deutsch als Muttersprache und gaben ihr mündliches Einverständnis zur Teilnahme. Als Aufwandsentschädigung erhielten die Versuchspersonen entweder Versuchspersonenstunden oder eine monetäre Vergütung.

Material. Als Lernmaterial wurden 24 schwach miteinander verwandte Wortpaare (z. B. Akrobat - Seil) verwendet, die aus den Normen von Nelson et al. (1998) entnommen und ins Deutsche übersetzt wurden. Die vorwärtsgerichtete Assoziationsstärke der Wortpaare lag dabei zwischen .01 und .07 und betrug im Mittel einen Wert von .03. Das bedeutet, dass in der Ursprungsstudie von Nelson et al. (1998) die Versuchspersonen in etwa 3 % der Fälle das korrekte Target-Wort als erste Antwort gaben, wenn ihnen ein Cue-Wort präsentiert wurde. Diese 24 Wortpaare in 2 Untergruppen (Untergruppe A und Untergruppe B) mit jeweils 12 Wortpaaren unterteilt. Für die Hälfte der Versuchspersonen wurde Untergruppe A in Vortest-Trials und Untergruppe B in Lern-Trials verwendet (alle geraden

Versuchspersonen-Nummern); für die andere Hälfte der Versuchspersonen war diese Zuordnung genau umgekehrt (alle ungeraden Versuchspersonen-Nummern).

Wie in der Studie von Carpenter (2011) bestand der finale Rekognitionstest aus insgesamt 96 Items, von denen 48 bereits bekannte, alte Items (jeweils 24 Cue-Wörter und Target-Wörter der gelernten Wortpaare) und 48 unbekannte, neue Items waren. Dabei waren 24 dieser unbekannten Items semantische Mediatoren. Genauer wurden für jedes der 24 Wortpaare zwei semantische Mediatoren, d. h. Wörter mit einer starken vorbestehenden Assoziation zu dem Cue-Wort, ebenfalls aus den Normen von Nelson et al. (1998) entnommen, wodurch sich eine Gesamtanzahl von 48 semantischen Mediatoren ergab. Diese 48 Wörter wurden dann in zwei Listen, eine Standard-Liste und eine Ausweich-Liste, mit jeweils 24 Wörtern unterteilt, die jeweils einen Mediator für jedes der 24 Wortpaare beinhaltete. Beispielsweise wurden für das Cue-Wort *Akrobat*, die Items *Turner* und *Zirkus* als semantische Mediatoren verwendet, die eine vorwärtsgerichtete Assoziationsstärke von .20 bzw. .17 zum Cue-Wort, aber keine bestehende Assoziation zum Target-Item *Seil* hatten. Die mittlere vorwärtsgerichtete Assoziationsstärke zwischen dem Cue-Item und den entsprechenden Mediatoren war für die Standard-Liste und die Ausweich-Liste identisch (jeweils .24). Im finalen Rekognitionstest wurde dann standardmäßig der Mediator der Standardliste als neues unbekanntes Item verwendet, es sei denn, dieser Mediator wurde von der Versuchsperson als Rateversuch im Vortest generiert. Für den Fall, dass der Mediator der Standard-Liste im initialen Vortest als Rateversuch genannt wurde, wurde im finalen Rekognitionstest der Mediator der Ausweichliste für dieses entsprechende Cue-Wort verwendet¹.

Die übrigen 24 neuen Items bestanden aus Wörtern, die komplett unverwandt zu allen Cue-Wörtern, Target-Wörtern und Mediatoren waren, und ebenfalls den Normen von Nelson et al. (1998) entnommen wurden. Die mittlere Wortlänge unterschied sich nicht signifikant

¹ Um zu gewährleisten, dass die mittlere vorwärtsgerichtete Assoziationsstärke zwischen einem Cue-Wort und dem entsprechenden Mediator über alle Versuchspersonen hinweg für die Lern-Bedingung und die Vortest-Bedingung vergleichbar war, wurde folgendes Verfahren angewandt: Wenn die Versuchsperson X den Mediator der Standard-Liste (z. B. *Turner*) im initialen Vortest als Rateversuch abgegeben hat, wurde wie oben beschrieben im finalen Test der Mediator der Ausweich-Liste (z. B. *Zirkus*) verwendet. Für die nachfolgende Versuchsperson (X+1), wurde derselbe Mediator dann auch als Mediator der Lern-Bedingung verwendet, sodass über alle Versuchspersonen hinweg alle Mediator-Wörter gleich häufig als Mediatoren der Lern-Bedingung und der Vortest-Bedingung verwendet wurden.

zwischen den Cue-Wörtern, den Target-Wörtern, den unverwandten Wörtern und den Mediatoren der Standard-Liste und der Ausweich-Liste (mittlere Anzahl an Buchstaben: 7.3, 5.8, 6.7, 6.1, 7.1), $F(4,115) = 1.95$, $MSE = 5.03$, $p = .11$, $\eta^2_p = .06$. Die Auftretenshäufigkeit pro eine Millionen Wörter gemäß der *Hyperspace Analogue to Language frequency scale* (Balota et al., 2007) unterschied sich ebenfalls nicht signifikant zwischen den fünf Item-Arten (Häufigkeit pro eine Million: 236.2, 326.7, 85.6, 406.5, 235.3), $F(4,115) = 2.24$, $MSE = 154213.1$, $p = .07$, $\eta^2_p = .07$. Alle 96 Wörter des finalen Tests wurden ins Deutsche übersetzt.

Design und Versuchsablauf. Das gesamte Experiment wurde online mittels der Videotelefonie-Software Zoom (Zoom Video Communications Inc., 2016) durchgeführt, wobei alle Testungen in einzelnen Videokonferenzen stattfanden. Das Experiment bestand dabei aus drei Phasen, die von jeder Versuchsperson durchlaufen wurden: einer Lern-Phase, einer Distraktor-Phase und einer finalen Test-Phase. Während der Lern-Phase wurden den Versuchspersonen alle 24 Wortpaare präsentiert, wobei die Hälfte der Cue-Wörter jeweils für 10 s zusammen mit dem entsprechenden Target-Wort (z. B. Akrobat – *Seil*) präsentiert wurden (Lern-Bedingung). Bei der anderen Hälfte der Items wurde für 5 s zunächst nur das Cue-Wort präsentiert und die Versuchspersonen wurden aufgefordert das Target-Wort mündlich zu erraten (z. B. Scheidung - ?), bevor das vollständige Wortpaar (z. B. Scheidung – *Anwalt*) für weitere 5 s auf dem Bildschirm zu sehen war (Vortest-Bedingung). In beiden Bedingungen wurden die Versuchspersonen anschließend dazu aufgefordert auf einer 5-stufigen Skala von 1 (*überhaupt kein Zusammenhang*) bis 5 (*sehr starker Zusammenhang*) zu bewerten, wie stark der subjektiv eingeschätzte Zusammenhang zwischen dem Cue-Wort und dem Target-Wort war. Die Versuchspersonen gaben ihre Antwort auch hier mündlich ab und es gab kein Zeitlimit, jedoch wurden die Versuchspersonen gebeten, möglichst intuitiv zu antworten. Um einen möglichen Deckeneffekt zu vermeiden, wurden die Versuchspersonen nicht darüber informiert, dass die Wortpaare gelernt und in einem späteren finalen Test abgerufen werden sollen. Stattdessen diente die Aufgabe zur Einschätzung des Zusammenhangs zwischen den Cue-Wörtern und den Target-Wörtern als implizite Erklärung, warum die Wortpaare präsentiert wurden. Alle Wortpaare wurden für jede Versuchsperson in einer zufälligen Reihenfolge dargeboten.

In der nachfolgenden Distraktor-Phase wurden die Versuchspersonen dazu aufgefordert, für 5 min ein Online-Bilderrätsel, bei dem Unterschiede zwischen zwei Bildern gesucht werden sollten, zu spielen (<https://www.suchbilder.com/fehlerbilder/>), für 12 min Ravens progressive Matrizen-Aufgaben (Raven et al., 2000) zu lösen und für 3 min Tetris zu spielen (<https://www.geo.de/geolino/spiele/13349-rtklonlinespiel-tetris>). Abschließend folgte der finale Rekognitionstest, in dem insgesamt 96 Wörter, bestehend aus 48 alten Wörtern (jeweils die 24 Cue- und Target-Wörter der Lern-Phase) und 48 neuen Wörtern (24 semantische Mediatoren und 24 unverwandte Wörter) in einer zufälligen Reihenfolge präsentiert wurden. Jedes Wort wurde laut vom Versuchsleiter vorgelesen und die Aufgabe der Versuchspersonen bestand darin, zu beurteilen, ob das präsentierte Wort ein altes, aus der Lern-Phase bekanntes Wort oder ein neues unbekanntes Wort war. Dabei war während des gesamten Tests eine 6-stufige Skala von 1 (*sicher neu*) bis 6 (*sicher alt*) auf dem Bildschirm zu sehen und die Versuchspersonen gaben ihre Antwort, indem sie die entsprechende Zahl laut sagten. Dabei gab es weder ein Zeitlimit für die Abgabe der Antwort noch ein Feedback über die Korrektheit der Antwort. Insgesamt ergab sich folglich für Experiment 5 ein 2 x 4 Design mit den Innersubjekt-Faktoren ÜBUNG (Lern-Bedingung vs. Vortest-Bedingung) und ITEM TYP (Target-Item vs. Cue-Item vs. Mediator-Item vs. Unverwandtes-Item).

Ergebnisse

Initialer Vortest. Während des initialen Vortests gelang es den Versuchspersonen in 96.5 % der Trials einen Rateversuch innerhalb der vorgegebenen 5 s abzugeben. Dabei wurden im Durchschnitt 4.6 % der Target-Wörter korrekt erraten. Alle Wortpaare, für die im Vortest das korrekte Target-Wort erraten wurde, wurden von den nachfolgenden Analysen ausgeschlossen, da insbesondere der Effekt falscher Rateversuche auf die nachfolgende Gedächtnisleistung von Interesse war.

Finaler Test.

Hit Rate. Dem Vorgehen von Carpenter (2011) folgend, wurde die Erinnerungsleistung im finalen Rekognitionstest getrennt für alte Items und neue Items jeweils in Abhängigkeit des Faktors ÜBUNG (Lern-Bedingung vs. Vortest-Bedingung) analysiert. Alle abgegebenen Antworten, die einem Zahlenwert von 4 oder höher

entsprachen, wurden als „Ja“ Antworten gewertet und entsprachen somit einem Hit für ein altes Item, da die Versuchsperson ein altes Item korrekt als altes Item eingeordnet hat, und einem False Alarm für ein neues Item, da ein neues Item fälschlicherweise als alt eingeordnet wurde. Alle Analysen wurden dabei für alle „Ja“ Antworten (Zahlenwerte von 4, 5 und 6), für „Ja“ Antworten mit mittlerer Antwortsicherheit (Zahlenwerte von 5 und 6) und für „Ja“ Antworten mit einer hohen Antwortsicherheit (Nur Zahlenwert 6) durchgeführt. Tabelle 1 zeigt die Hit Raten für die unterschiedlichen Niveaus an Antwortsicherheit in Abhängigkeit der Faktoren ÜBUNG (Lern-Bedingung vs. Vortest-Bedingung) und ITEM-TYP (Cue-Item vs. Target-Item). Da die Cue-Items im finalen Test vor allem aus dem Grund verwendet wurden, eine identische Anzahl an alten und neuen Items zu gewährleisten, und das Hauptinteresse der Hit Rate der Target-Items galt, wurden die Hit Raten der Cue-Items getrennt und nicht als Teil einer ANOVA untersucht (siehe auch Carpenter, 2011).

Tabelle 1

Hit Rate in Experiment 5

Niveau der Antwortsicherheit	Lernen	Vortest
<i>Alle „Ja“ Antworten (4er, 5er, 6er)</i>		
Cue-Items	89.51 (1.59)	96.11 (0.94)
Target-Items	86.57 (1.58)	90.01 (1.51)
<i>„Ja“ Antworten mit mittlerer Sicherheit (5er, 6er)</i>		
Cue-Items	88.12 (1.80)	94.79 (1.17)
Target-Items	82.72 (1.98)	86.51 (1.84)
<i>„Ja“ Antworten mit hoher Sicherheit (nur 6er)</i>		
Cue-Items	81.94 (2.44)	90.52 (1.75)
Target-Items	76.70 (2.32)	79.40 (2.74)

Anmerkung: Hit Rate (in %) der alten Items (Cue-Items und Target-Items) für die unterschiedlichen Niveaus an Antwortsicherheit in Abhängigkeit des Faktors Übung (Lernen vs. Vortest). In den Klammern befindet sich der Standardfehler des Mittelwertes.

Geplante paarweise Vergleiche offenbarten einen signifikanten Vortest-Effekt, der sich in einer höheren Hit Rate für Target-Items der Vortest-Bedingung verglichen mit der Lern-Bedingung zeigte, sowohl bei Betrachtung aller „Ja“ Antworten (0.90 vs. 0.87), $t(53) = 2.26$, $p = .028$, Cohen's $d = 0.31$, als auch bei Betrachtung der „Ja“ Antworten mit mittlerer Antwortsicherheit (0.87 vs. 0.83), $t(53) = 2.39$, $p = .020$, Cohen's $d = 0.33$. Es zeigte sich jedoch kein signifikanter Unterschied bei der Betrachtung der „Ja“ Antworten mit hoher Antwortsicherheit (0.79 vs. 0.77), $t(53) = 1.50$, $p = .14$, Cohen's $d = 0.20$. Für die Cue-Items zeigte sich ein signifikanter Vortest-Effekt unabhängig vom Niveau der Antwortsicherheit, alle $t_s > 3.89$, alle $p_s < .001$.

Tabelle 2

False Alarms Rate in Experiment 5

Niveau der Antwortsicherheit	Lernen	Vortest
<i>Alle „Ja“ Antworten (4er, 5er, 6er)</i>		
Mediator-Items	4.17 (0.85)	8.39 (1.33)
Unverwandte Items	2.47 (0.68)	1.70 (0.64)
<i>„Ja“ Antworten mit mittlerer Sicherheit (5er, 6er)</i>		
Mediator-Items	2.62 (0.62)	4.79 (0.87)
Unverwandte Items	1.70 (0.56)	0.93 (0.42)
<i>„Ja“ Antworten mit hoher Sicherheit (nur 6er)</i>		
Mediator-Items	0.77 (0.40)	1.83 (0.50)
Unverwandte Items	1.23 (0.46)	0.31 (0.22)

Anmerkung: False Alarms Rate (in %) für die neuen Items (Mediator-Items und unverwandte Items) für die unterschiedlichen Niveaus an Antwortsicherheit in Abhängigkeit des Faktors Übung (Lernen vs. Vortest). In den Klammern befindet sich der Standardfehler des Mittelwertes.

False Alarm Rate. Tabelle 2 zeigt die False Alarm Rate der unterschiedlichen Niveaus an Antwortsicherheit in Abhängigkeit der Faktoren ÜBUNG (Lern-Bedingung vs. Vortest-Bedingung) und ITEM-TYP (Mediator-Item vs. unverwandtes Item). Eine 2 x 2

ANOVA der beiden Innersubjekt-Faktoren bezüglich der False Alarm Rate aller „Ja“ Antworten zeigte einen signifikanten Haupteffekt ITEM-TYP, $F(1, 53) = 19.74$, $MSE = 25.48$, $p < .001$, $\eta^2_p = .27$, mit einer höheren False Alarm Rate für Mediator-Items als für unverwandte Items (6.3 % vs. 2.1 %), aber keinen signifikanten Haupteffekt ÜBUNG, $F(1, 53) = 3.13$, $MSE = 25.48$, $p = .083$, $\eta^2_p = .06$. Vor allem aber zeigte sich eine signifikante Interaktion zwischen den beiden Faktoren, $F(1, 53) = 13.20$, $MSE = 25.48$, $p < .001$, $\eta^2_p = .20$. Geplante paarweise Vergleiche offenbarten eine höhere False Alarm Rate in der Vortest-Bedingung verglichen mit der Lern-Bedingung für Mediator-Items (8.4 % vs. 4.2 %), $t(53) = 3.06$, $p = .004$, Cohen's $d = 0.42$, aber keinen signifikanten Unterschied zwischen den beiden Bedingungen für die unverwandten Items (1.7 % vs. 2.5 %), $t(53) = 0.80$, $p = .43$, Cohen's $d = 0.11$. Sowohl bei der Betrachtung der „Ja“ Antworten mit mittlerer Antwortsicherheit als auch bei Betrachtung der „Ja“ Antworten mit hoher Antwortsicherheit zeigte sich das gleiche Ergebnismuster mit jeweils einer signifikanten Interaktion der beiden Faktoren, alle $F_s > 6.53$ und allen $p_s < .013$.

Zusätzliche Analyse. Wie in der Vorgängerstudie von Carpenter (2011) wurde auch im vorliegenden Experiment 1 die Leistung im finalen Rekognitionstest jeweils getrennt für die Hit Rate und die False Alarm Rate analysiert. Dieses Vorgehen bezüglich der Datenauswertung wurde gewählt, um die Ergebnisse des vorliegenden Experiments so gut wie möglich mit den Befunden der Vorgängerstudie vergleichen zu können. Da sich in den Ergebnissen sowohl eine höhere Hit Rate für die alten Items als auch eine höhere False Alarm Rate für die Mediator-Items in der Vortest-Bedingung verglichen mit der Lern-Bedingung zeigte, stellt sich die Frage, ob dieses Ergebnismuster durch ein in der Vortest-Bedingung liberaleres Antwortkriterium als in der Lern-Bedingung zustande gekommen ist. Dieses liberalere Antwortkriterium könnte sich in einer grundsätzlich erhöhten Wahrscheinlichkeit zeigen, Items als alt einzuordnen, was einerseits zu einer höheren Hit Rate für tatsächliche alte Target-Items (und Cue-Items) und andererseits zu einer erhöhten False Alarm Rate für die Mediator-Items führen sollte. Um diese Möglichkeit zu untersuchen, wurde eine ROC (Receiver Operating Characteristic) Analyse durchgeführt, mit der der mögliche Antwortbias in Lern-Trials und in Vortest-Trials genauer betrachtet werden sollten. Dazu wurde der Anteil an korrekt wiedererkannten Target-Items (die Hit Rate) und der Anteil an fehlerhaft als alt eingeschätzten neuen Items (False Alarm Rate) für die gesamte Rating-Skala

kumuliert, beginnend mit dem sichersten Antwortkriterium (definitiv alt, „6“). Diese Methode erzeugt eine empirische ROC-Kurve, die den Zusammenhang zwischen Hits und False Alarms für die unterschiedlichen Antwortkriterien abbildet (z. B. Macmillan & Creelman, 2004). Durch die Verwendung einer 6-stufigen Skala wurden die Hit Rate und die False Alarm Rate somit für fünf unterschiedliche Antwortkriterien ermittelt.

Im nächsten Schritt wurden die Daten des finalen Rekognitionstests mittels des Ansatzes der Signalentdeckungstheorie untersucht. Dabei wurde eine ungleiche Varianz für die Verteilung der alten und neuen Items angenommen, um die typische asymmetrische Form der ROC zu erfassen, und die Daten wurden mittels des Signalentdeckungs-Modells der ungleichen Varianz (z. B. Wixted, 2007) beschrieben. Dieses Modell stützt die abgegebenen Einschätzungen im Rekognitionstest auf eine einzelne Information, nämlich die gesamt betrachtete Gedächtnisstärke der Items. Überschreitet ein Item das Antwortkriterium c_i , das mit einem bestimmten Konfidenzniveau i verbunden ist, aber nicht das Kriterium c_{i-1} , bewerten die Teilnehmer das Item entsprechend mit i . Die Erinnerungsstärke alter Items im Vergleich zu neuen Items - also die Sensitivität - lässt sich aus dem Abstand zwischen den Mittelwerten der zugrundeliegenden Stärkeverteilungen der alten und neuen Items ableiten (d'). Die Anwendung des Modells auf die 5-Punkte-ROC-Daten führt zu sieben freien Parametern (Gedächtnisstärke der alten Items d_a , Varianz der Verteilung der alten Items λ und fünf Kriteriumspunkte c_1 - c_5) und folglich zu drei Freiheitsgraden beim Testen der Anpassungsgüte des Modells. Zur Schätzung der Modellparameter wurde die Maximum-Likelihood-Methode verwendet, die auch für statistische Tests eingesetzt werden kann.

Tabelle 3 zeigt die Statistik der Anpassungsgüte sowie die Maximum-Likelihood-Schätzungen der Modellparameter d' , λ und c_1 bis c_5 sowohl für die Lern-Trials als auch für die Vortest-Trials. Das Modell der Signalentdeckung mit ungleicher Varianz lieferte eine gute Anpassung an die Rekognitionsdaten der beiden Übungsarten (Lern-Bedingung vs. Vortest-Bedingung), $\chi^2(3) \leq 1.83$, alle $ps \geq 0.59$. Weder bezüglich des Sensitivitätsmaßes d' noch bezüglich λ wurde der numerische Unterschied zwischen den Lern-Trials und den Vortest-Trials signifikant, alle $\chi^2(1) \leq 0.56$, alle $ps \geq 0.45$. Der Befund, dass die Sensitivität für die Lern-Bedingung und die Vortest-Bedingung ähnlich war, stimmt mit der oben beschriebenen Beobachtung überein, dass Vortests sowohl die Hit Rate der Target-Items als auch die False Alarm Rate der Mediator-Items erhöhen. Die statistische Analyse der

Modellparameter zeigte des Weiteren, dass die geschätzten Antwortkriterien c_1 - c_4 in der Vortest-Bedingung signifikant reduziert - d. h. liberaler - waren als in der Lern-Bedingung, alle $\chi^2s(1) \geq 6.76$, alle $ps \leq 0.009$. Diese Analyse deutet darauf hin, dass die Versuchspersonen in der Vortest-Bedingung eine stärkere Tendenz zeigten, Items als alt zu kategorisieren, unabhängig davon, ob das Item während der Lern-Phase tatsächlich präsentiert wurde.

Tabelle 3

Anpassungsgüte des Modells mit ungleicher Varianz für die Rekognitionsdaten

Bedingung	Parameterschätzungen							Anpassungsgüte		
	d'	λ	c_1	c_2	c_3	c_4	c_5	χ^2	df	p
Lernen	3.73	1.88	2.34	1.91	1.70	1.25	0.47	3.00	3	0.39
Vortest	3.62	1.77	2.10	1.64	1.39	1.03	0.45	1.38	3	0.71

Anmerkung: d' = allgemeine Gedächtnisstärke; λ = Varianz der Verteilung der alten Items; c_{1-5} = Antwortkriterien; df = Freiheitsgrade.

Diskussion

Die Ergebnisse von Experiment 5 konnten für die alten Items den klassischen Vortest-Effekt zeigen, der sich in einer höheren Hit Rate für Target-Items der Vortest-Bedingung verglichen mit Target-Items der Lern-Bedingung widerspiegelt und somit bisherige Studien, die unter Verwendung eines Rekognitionstests ebenfalls einen Erinnerungsvorteil von Vortest zeigen konnten, replizieren (z. B. Seabrooke et al., 2019). Noch wichtiger war jedoch der Befund, dass zudem auch eine höhere False Alarm Rate für neue Mediator-Items, jedoch nicht für neue, unverwandte Items in der Vortest-Bedingung verglichen mit der Lern-Bedingung gefunden werden konnte. Die vorliegenden Befunde generalisieren damit die von Carpenter (2011, Experiment 1) gefundenen Ergebnisse vom klassischen Test-Effekt-Paradigma auf das Vortest-Paradigma. Jedoch besteht ein Unterschied zwischen den vorliegenden Ergebnissen und den Ergebnissen von Carpenter darin, dass in der Vorgängerstudie der klassische Test-Effekt am größten für Antworten mit hoher Antwortsicherheit ausfiel, wohingegen in der vorliegenden Studie das Niveau der

Antwortsicherheit keinen Einfluss auf die Größe des Vortest-Effekts zu haben schien. Dieser Unterschied lässt sich nicht notwendigerweise nur darauf zurückführen, dass in der Vorgängerstudie ein Test-Paradigma und in der aktuellen Studie ein Vortest-Paradigma verwendet wurde, sondern ist möglicherweise auch durch grundsätzliche prozedurale Unterschiede (z. B. wurden den Versuchspersonen bei Carpenter weniger Wortpaare in der Lern-Phase präsentiert und die Versuchspersonen wurden über das Stattfinden eines finalen Tests informiert) zu erklären.

Die vorliegenden Ergebnisse stützen die Elaborations-Hypothese des Vortests dadurch, dass hier erstmalig gezeigt wurde, dass durch Vortests die Wahrscheinlichkeit der Aktivierung semantischer Mediatoren erhöht zu werden scheint. Genauer betrachtet scheint die Aktivierung semantischer Mediatoren während des initialen Vortests nicht nur die Gedächtnisrepräsentation der Target-Items durch eine verbesserte Integration der Target-Items mit den Cue-Items zu stärken, sondern auch die Gedächtnisrepräsentation der Mediator-Items zu verbessern. Durch die Stärkung sowohl der Target-Items als auch der Mediator-Items könnte bei den Versuchspersonen häufiger der Eindruck entstehen, dass diese Items bereits zuvor präsentiert wurden, was zu einem liberaleren Antwortbias in der Vortest-Bedingung verglichen mit der Lern-Bedingung geführt haben könnte (siehe auch Spitzer & Bäuml, 2009, die einen ähnlichen Effekt von Anrufübung auf den Antwortbias berichten). Das Ziel von Experiment 6 war es, zu untersuchen, ob semantische Mediatoren nicht nur mit einer höheren Wahrscheinlichkeit während eines Vortests aktiviert werden, sondern in der Vortest-Bedingung auch mit höherer Wahrscheinlichkeit mit dem Target-Item verknüpft werden als in der Lern-Bedingung. Um diese Vorhersage zu testen, sollte in Experiment 6 betrachtet werden, ob die Präsentation eines semantischen Mediators als Cue-Item in einem finalen Test zu einer besseren Erinnerung an ein Target-Item führt, wenn für die Items zuvor ein Vortest bearbeitet wurde, verglichen mit dem bloßen Lernen der Items.

7.2. Experiment 6

Method

Stichprobe. Der Poweranalyse aus Experiment 5 folgend, nahmen für Experiment 6 jeweils 40 Versuchspersonen in jeder der beiden Zwischensubjekt-Bedingungen teil,

wodurch sich eine Gesamtstichprobe von $N = 80$ Versuchspersonen ($M = 22.8$ Jahre, Altersspanne = 18-34 Jahre, 67 weiblich, 12 männlich, 1 divers) ergab. Alle Versuchspersonen sprachen Deutsch als Muttersprache und gaben ihr mündliches Einverständnis zur Teilnahme. Als Aufwandsentschädigung erhielten die Versuchspersonen erneut entweder Versuchspersonenstunden oder eine monetäre Vergütung.

Material. Sowohl das verwendete Material als auch die Aufteilung des Materials auf die verschiedenen Versuchsbedingungen war identisch zu Experiment 5.

Design und Versuchsablauf. Auch der Versuchsablauf war identisch zu Experiment 5, mit der einzigen Ausnahme, dass der finale Test statt aus einem Rekognitionstest aus einem Cued-Recall-Test bestand. Zu Beginn dieses finalen Cued-Recall-Tests wurden alle Versuchspersonen darüber informiert, dass der Versuchsleiter im Anschluss Wörter laut vorlesen wird, die entweder einem Cue-Wort aus der Lern-Phase oder einem neuen Wort entsprechen. Die Aufgabe der Versuchspersonen bestand darin, immer dann, wenn sie durch ein genanntes Wort an ein Target-Wort aus der Lern-Phase erinnert wurden (z. B. *Seil*), dieses Target-Wort laut zu nennen, unabhängig davon, ob das vorgelesene Wort ein bekanntes oder ein neues Wort darstellte. Für die Hälfte der Versuchspersonen wurden ausschließlich die ursprünglich präsentierten (originalen) Cue-Items (z. B. Akrobat) als Cue-Wörter im finalen Test verwendet, wohingegen die andere Hälfte der Versuchspersonen ausschließlich Mediator-Items (z. B. Turner) als Cue-Wörter im finalen Test präsentiert bekam. Für Experiment 6 ergab sich folglich ein 2×2 Design mit dem Innersubjektfaktor ÜBUNG (Lern-Bedingung vs. Vortest-Bedingung) und dem Zwischensubjektfaktor CUE-TYP (originaler Cue vs. Mediator-Cue). Während des finalen Tests gab es weder ein Zeitlimit noch korrekatives Feedback.

Ergebnisse

Initialer Vortest. Während des initialen Vortests gelang es den Versuchspersonen in 98.7 % der Trials einen Rateversuch innerhalb des 5 s Zeitrahmens abzugeben. Dabei wurde im Mittel für 4.6 % der Wortpaare das korrekte Target-Item erraten. Alle Wortpaare, für die im Vortest das korrekte Target-Wort erraten wurde, wurden von den nachfolgenden Analysen ausgeschlossen, da das Hauptinteresse dem Effekt falscher Rateversuche auf die nachfolgende Gedächtnisleistung galt.

Finaler Test.

Korrektur Abruf. Tabelle 4 zeigt den prozentualen Anteil an korrekt erinnerten Target-Items im finalen Cued-Recall-Test in Abhängigkeit der Faktoren ÜBUNG (Lern-Bedingung vs. Vortest-Bedingung) und CUE-TYP (originaler Cue vs. Mediator-Cue). Eine 2 x 2 ANOVA der beiden Faktoren ergab dabei sowohl einen signifikante Haupteffekt ÜBUNG, $F(1,78) = 19.66$, $MSE = 111.15$, $p < .001$, $\eta^2_p = .20$, als auch einen signifikanten Haupteffekt CUE-TYP, $F(1,78) = 592.95$, $MSE = 111.15$, $p < .001$, $\eta^2_p = .88$, mit einer insgesamt besseren Erinnerungsleistung für Target-Items in der Vortest-Bedingung verglichen mit der Lern-Bedingung (68.8 % vs. 61.5 %) und einer Abnahme der Erinnerungsleistung, wenn im finalen Test Mediator-Cues statt der originalen Cues präsentiert wurden (39.5 % vs. 90.8 %). Vor allem zeigte sich auch eine signifikante Interaktion zwischen den beiden Faktoren, $F(1,78) = 17.83$, $MSE = 111.15$, $p < .001$, $\eta^2_p = .19$. Geplante paarweise Vergleiche offenbarten keinen signifikanten Unterschied in der Erinnerungsleistung zwischen der Lern-Bedingung und der Vortest-Bedingung, wenn der originale Cue präsentiert wurde (91.0 % vs. 90.6 %), $t(39) = 0.19$, $p = .85$, Cohen's $d = 0.03$, und somit keinen Vortest-Effekt. Wurde hingegen im finalen Test der Mediator-Cue verwendet, war die Erinnerungsleistung für die Target-Items in der Vortest-Bedingung signifikant besser als in der Lern-Bedingung (46.7 % vs. 32.3 %), $t(39) = 5.26$, $p < .001$, Cohen's $d = 0.83$, womit sich ein Vortest-Effekt von 14.4 % für diese Art des Cues zeigte.

Tabelle 4

Korrektur Abruf in Experiment 6

		Übung	
		Lernen	Vortest
Cue-Typ	Originaler Cue	90.63 (1.27)	90.98 (1.65)
	Mediator-Cue	32.29 (2.07)	46.72 (2.41)

Anmerkung: Erinnerungsleistung (in %) in Abhängigkeit der Faktoren Übung (Lernen vs. Vortest) und Cue-Typ (originaler Cue vs. Mediator-Cue). In den Klammern befindet sich der Standardfehler des Mittelwertes.

Intrusionen. Alle fehlerhaften Antworten, die im finalen Cued-Recall-Test von den Versuchspersonen produziert wurden, wurden als Intrusionen gewertet. Die mittlere Anzahl an Intrusionen in Abhängigkeit der Faktoren ÜBUNG (Lern-Bedingung vs. Vortest-Bedingung) und CUE-TYP (originaler Cue vs. Mediator-Cue) ist in Tabelle 5 abgebildet. Eine 2 x 2 ANOVA der beiden Faktoren zeigte einen signifikanten Haupteffekt CUE-TYP, $F(1,78) = 41.36$, $MSE = 2.36$, $p < .001$, $\eta^2_p = .35$, mit einer insgesamt niedrigeren Anzahl an Intrusionen, wenn ein originaler Cue im finalen Test präsentiert wurde verglichen mit einem Mediator-Cue (0.4 Intrusionen vs. 2.0 Intrusionen). Es zeigte sich jedoch weder ein signifikanter Haupteffekt ÜBUNG, $F(1,78) = 3.82$, $MSE = 2.36$, $p = .054$, $\eta^2_p = .05$, noch eine signifikante Interaktion der beiden Faktoren, $F(1,78) = 1.51$, $MSE = 2.36$, $p = .22$, $\eta^2_p = .02$.

Tabelle 5*Intrusionen in Experiment 6*

		Übung	
		Lernen	Vortest
Cue-Typ	Originaler Cue	0.50 (0.11)	0.38 (0.13)
	Mediator-Cue	2.23 (0.31)	1.73 (0.23)

Anmerkung: Mittlere Anzahl an Intrusionen im finalen Test in Abhängigkeit der Faktoren Übung (Lernen vs. Vortest) und Cue-Typ (originaler Cue vs. Mediator-Cue). In den Klammern befindet sich der Standardfehler des Mittelwertes.

Diskussion

Die Ergebnisse von Experiment 6 liefern erste Hinweise darauf, dass durch das Bearbeiten eines Vortests semantische Mediatoren nicht nur mit einer höheren Wahrscheinlichkeit aktiviert werden, sondern in der Vortest-Bedingung auch mit höherer Wahrscheinlichkeit mit dem Target-Item verknüpft werden als in der Lern-Bedingung. So zeigte sich in Experiment 6 ein signifikanter Vortest-Effekt, wenn im finalen Test ein semantischer Mediator als Cue verwendet wurde, was sich in einer besseren

Erinnerungsleistung an das Target-Item in der Vortest-Bedingung verglichen mit der Lern-Bedingung widerspiegelte. Wie bei Experiment 5 generalisieren auch die Befunde von Experiment 6 die Ergebnisse von Carpenter (2011, Experiment 2) vom klassischen Test-Paradigma auf das Vortest-Paradigma. Allerdings blieb im vorliegenden Experiment der typische Vortest-Effekt unter Verwendung der originalen Cues im finalen Test aus. Da jedoch die Erinnerungsquote in der Lern-Bedingung mit 91.0 % bereits sehr hoch war, gab es aufgrund dieses Deckeneffekts nicht viel Spielraum für eine Verbesserung der Erinnerungsleistung in der Vortest-Bedingung, was möglicherweise den ausbleibenden Vortest-Effekt erklären könnte. Um zu untersuchen, inwiefern der ausbleibende Vortest-Effekt tatsächlich auf einen potenziellen Deckeneffekt zurückzuführen ist und ob sich die Befunde bezüglich der Mediator-Cues auch dann replizieren lassen, wenn der typische Vortest-Effekt vorhanden ist, wurde ein weiteres Experiment 7 durchgeführt. Dazu wurde dieselbe Aufgabe wie in Experiment 6 verwendet mit dem Unterschied, dass die Schwierigkeit des finalen Tests durch eine Verlängerung des Retentionsintervalls von 20 min auf 4 h erhöht wurde, um die Erinnerungsquoten der Lern-Bedingung zu reduzieren und somit mehr Spielraum für einen potenziellen Vorteil von Vortests auf die spätere Erinnerungsleistung zu ermöglichen.

7.3. Experiment 7

Methode

Stichprobe. Wie in Experiment 6 nahmen auch an Experiment 7 insgesamt 80 Versuchspersonen teil ($M = 24.1$ Jahre, Altersspanne = 20-35 Jahre, 47 weiblich), wobei wieder jeweils 40 Versuchspersonen zufällig den beiden Zwischensubjektbedingungen zugeteilt wurden. Alle Versuchspersonen sprachen Deutsch als Muttersprache und gaben ihr mündliches Einverständnis zur Teilnahme. Als Aufwandsentschädigung erhielten die Versuchspersonen erneut entweder Versuchspersonenstunden oder eine monetäre Vergütung.

Material., Design und Versuchsablauf. Sowohl das Material als auch das Design und der Versuchsablauf waren identische zu Experiment 6 mit der einzigen Ausnahme des Retentionsintervalls zwischen der Lern-Phase und dem finalen Test. In Experiment 6 fand

der finale Cued-Recall-Test nach einer 20-minütigen Distraktorphase statt und die Testung konnte so innerhalb eines Termins beendet werden. In Experiment 7 hingegen war nach der Lern-Phase Teil 1 der Testung beendet und die Versuchspersonen kehrten 4 h nach der Lern-Phase für Teil 2 der Testung, in der der finale Cued-Recall-Test stattfand, zurück. Alle Versuchspersonen, die an Teil 1 teilnahmen, erschienen auch für Teil 2 und beendeten selbigen.

Ergebnisse

Initialer Vortest. Im initialen Vortest gaben die Versuchspersonen in 95.1 % der Trials einen gültigen Rateversuch innerhalb des 5 s Zeitrahmens ab. Dabei wurde das korrekte Target-Wort in durchschnittlich 4.6 % der Trials erraten und erneut wurden alle Wortpaare, für die ein korrekter Rateversuch abgegeben wurde, aus den weiteren Analysen ausgeschlossen.

Tabelle 6

Korrektur Abruf in Experiment 7

		Übung	
		Lernen	Vortest
Cue-Typ	Originaler Cue	77.08 (3.22)	81.10 (2.92)
	Mediator-Cue	27.92 (2.28)	45.02 (2.35)

Anmerkung: Erinnerungsleistung (in %) in Abhängigkeit der Faktoren Übung (Lernen vs. Vortest) und Cue-Typ (originaler Cue vs. Mediator-Cue). In den Klammern befindet sich der Standardfehler des Mittelwertes.

Finaler Test.

Korrektur Abruf. Der prozentuale Anteil an korrekt erinnerten Target-Items im finalen Cued-Recall-Test in Abhängigkeit der Faktoren ÜBUNG (Lern-Bedingung vs. Vortest-Bedingung) und CUE-TYP (originaler Cue vs. Mediator-Cue) ist in Tabelle 6 dargestellt. Eine 2 x 2 ANOVA der beiden Faktoren zeigte signifikante Haupteffekte

ÜBUNG, $F(1,78) = 84.94$, $MSE = 52.49$, $p < .001$, $\eta^2_p = .52$, und CUE-TYP, $F(1,78) = 134.32$, $MSE = 541.01$, $p < .001$, $\eta^2_p = .63$. Dabei wurden in der Vortest-Bedingung mehr Target-Items korrekt abgerufen als in der Lern-Bedingung (63.1 % vs. 52.5 %) und die Erinnerungsleistung fiel deutlich schlechter aus, wenn im finalen Test der Mediator-Cue präsentiert wurde, verglichen mit der Präsentation des originalen Cues (36.5 % vs. 79.1 %). Vor allem aber konnte auch eine signifikante Interaktion der beiden Faktoren gefunden werden, $F(1,78) = 32.64$, $MSE = 52.49$, $p < .001$, $\eta^2_p = .30$. Geplante paarweise Vergleiche offenbarten dabei eine bessere Erinnerungsleistung an die Target-Items in der Vortest-Bedingung verglichen mit der Lern-Bedingung, sowohl, wenn der originale Cue präsentiert wurde (81.1 % vs. 77.1 %), $t(39) = 3.32$, $p = .002$, Cohen's $d = 0.53$, woraus sich ein Vortest-Effekt von 4.0 % ergab, als auch, wenn der Mediator-Cue verwendet wurde (45.0 % vs. 27.9%), $t(39) = 8.78$, $p < .001$, Cohen's $d = 1.39$, wobei der Vortest-Effekt sogar 17.1 % betrug.

Tabelle 7

Intrusionen in Experiment 7

		Übung	
		Lernen	Vortest
Cue-Typ	Originaler Cue	1.35 (0.25)	0.90 (0.23)
	Mediator-Cue	3.13 (0.27)	2.23 (0.28)

Anmerkung: Mittlere Anzahl an Intrusionen im finalen Test in Abhängigkeit der Faktoren Übung (Lernen vs. Vortest) und Cue-Typ (originaler Cue vs. Mediator-Cue). In den Klammern befindet sich der Standardfehler des Mittelwertes.

Intrusionen. Zur Betrachtung der Anzahl an Intrusionen wurde ebenfalls eine 2 x 2 ANOVA mit den Faktoren ÜBUNG (Lern-Bedingung vs. Vortest-Bedingung) und CUE-TYP (originaler Cue vs. Mediator-Cue) durchgeführt, wobei sich sowohl ein signifikanter Haupteffekt ÜBUNG, $F(1,78) = 15.98$, $MSE = 1.06$, $p < .001$, $\eta^2_p = .17$, mit einer größeren Anzahl an Intrusionen in der Lern-Bedingung verglichen mit der Vortest-Bedingung (2.2

Intrusionen vs. 1.6 Intrusionen) als auch ein signifikanter Haupteffekt CUE-TYP, $F(1,78) = 23.00$, $MSE = 1.06$, $p < .001$, $\eta^2_p = .23$, mit einer größeren Anzahl an Intrusionen bei der Präsentation des Mediator-Cues verglichen mit dem originalen Cue (2.7 Intrusionen vs. 1.1 Intrusionen) zeigte (siehe Tabelle 7). Die Interaktion der beiden Faktoren wurde nicht signifikant, $F(1,78) = 1.51$, $MSE = 1.06$, $p = .22$, $\eta^2_p = .02$.

Diskussion

Die Ergebnisse von Experiment 7 konnten erneut zeigen, dass die Verwendung eines semantischen Mediator-Items als Cue-Item im finalen Test zu einem signifikanten Vortest-Effekt führen kann. Im Gegensatz zu Experiment 6 konnte in Experiment 7 allerdings auch unter Verwendung der originalen Cue-Items im finalen Test ein Erinnerungsvorteil für die Vortest-Bedingung verglichen mit der Lern-Bedingung gefunden werden, was den typischen Vortest-Effekt abbildete. Tatsächlich führte die Verlängerung des Retentionsintervalls zwischen der Lern-Phase und dem finalen Test von 20 Minuten (Experiment 6) auf 4 Stunden (Experiment 7) zu einer Reduktion der Erinnerungsquote der Lern-Bedingung um ca. 14 %, wodurch sich mehr Spielraum für eine potenzielle Verbesserung der Erinnerungsleistung in der Vortest-Bedingung, verglichen mit der Lern-Bedingung, ergab. Die Befunde aus Experiment 7 passen erneut zur Elaborations-Hypothese des Vortest-Effekts und stützen somit die Annahme, dass es durch die Bearbeitung eines Vortests mit höherer Wahrscheinlichkeit zu einer Aktivierung zusätzlicher Informationen, wie beispielsweise semantischer Mediatoren, kommt und diese auch mit höherer Wahrscheinlichkeit mit dem Target-Item verknüpft werden, als dies in der Lern-Bedingung der Fall ist. Diese zusätzliche Verknüpfung wiederum scheint dann den Abruf des Target-Items in einem späteren finalen Test zu erleichtern.

Als Erklärung für das Ergebnismuster der Experimente 6 und 7 könnte wie in Experiment 5 erneut das Argument herangezogen werden, dass es durch die Bearbeitung eines Vortests zu einer Verschiebung des Antwortkriteriums hin zu einem liberaleren Antwort-Bias gekommen sein könnte, der den beobachteten positiven Effekt von Vortests unter Verwendung semantischer Mediatoren als Cue-Items im finalen Test bedingt haben könnte. Sollte dies der Fall gewesen sein, so wäre davon auszugehen, dass in der Vortest-Bedingung insgesamt mehr Antworten abgegeben worden sind als in der Lern-Bedingung, was einerseits zu einer höheren Rate an korrekt erinnerten Target-Items, aber auch zu einer

höheren Anzahl an fehlerhaften Antworten (Intrusionen) im finalen Test in der Vortest-Bedingung geführt haben sollte. Entgegen dieser Annahme offenbarten die Ergebnisse der Experimente 6 und 7 zwar eine höhere Rate an korrekt erinnerten Target-Items, aber auch eine *niedrigere* Anzahl an Intrusionen (in Experiment 6 numerisch weniger Intrusionen, in Experiment 7 statistisch signifikant weniger Intrusionen) in der Vortest-Bedingung verglichen mit der Lern-Bedingung. Diese Ergebnisse legen nahe, dass möglicherweise das Format des finalen Tests eine entscheidende Rolle dabei spielt, ob Vortests zu einem liberaleren Antwort-Bias führen als in der Lern-Bedingung, da sich diese Verschiebung für einen Rekognitionstest (Experiment 5), nicht aber für einen Cued-Recall-Test (Experimente 6 und 7), finden ließ. Unter welchen Voraussetzungen die Bearbeitung eines Vortests zu einer Verschiebung des Antwortkriteriums der Versuchspersonen führt, bleibt somit noch ungeklärt.

7.4. Diskussion Experimente 5-7

Die Ergebnisse der vorliegenden Experimente 5 und 7 konnten zeigen, dass die Bearbeitung eines Vortests zu einer besseren Erinnerungsleistung für die Target-Items, verglichen mit den Lern-Items, führt. Dies war der Fall unabhängig davon, ob ein Rekognitionstest oder ein Cued-Recall-Test als finaler Test verwendet wurde. Somit ließen sich bisherige Studien replizieren, die ebenfalls für beide Testformate einen Vortest-Effekt zeigen konnten (z. B. Kornell et al., 2009, Seabrooke et al., 2021). Des Weiteren stellen die Befunde der Experimente 5 und 7 eine Erweiterung der bisherigen Literatur zum Vortest-Effekt dar. So wurden die Versuchspersonen in den bisherigen Studien zur Untersuchung des Vortest-Effekts zu Beginn der Experimente instruiert, dass die nachfolgend präsentierten Inhalte gelernt und in einem abschließenden finalen Test abgerufen werden sollen, d. h. das Lernen fand bewusst und zielgerichtet statt, weshalb hier von intentionalem Lernen gesprochen werden kann. In den vorliegenden Experimenten 5-7 hingegen fand das Lernen inzidentell statt, d. h. ohne eine konkrete Lerninstruktion wurden die Wortpaare unbewusst durch die Beurteilung des Zusammenhanges zwischen dem Cue-Wort und dem Target-Wort gelernt. Die vorliegenden Befunde legen somit erstmalig die Vermutung nahe, dass sich ein Vorteil von Vortests auf die spätere Erinnerungsleistung auch dann zeigt, wenn der Lernprozess

nicht bewusst stattfindet. In Experiment 6 hingegen konnte der klassische Vortest-Effekt nicht gefunden werden, was wahrscheinlich auf einen Deckeneffekt zurückzuführen war.

Das Hauptinteresse der vorliegenden drei Experimente war jedoch die Frage, ob sich die von Carpenter (2011) für das klassische Test-Effekt-Paradigma gefundenen Ergebnismuster auch unter Verwendung des Vortest-Effekt-Paradigmas replizieren und somit auf dieses Paradigma generalisieren lassen. Hier liefern die Experimente 5-7 erste Hinweise darauf, dass sich für beide Paradigmen ein ähnliches Ergebnismuster zeigt, wenn semantische Mediatoren als Teil eines finalen Abruftestes verwendet werden. So führte in Experiment 5 das Bearbeiten eines Vortests, in dem die Target-Items erraten werden sollten (z. B. Mutter - ?), zu einer höheren False Alarms Rate für semantische Mediatoren (z. B. Vater) als in der Lern-Bedingung, d. h. mehr Items, die in der Lern-Phase nicht präsentiert wurden, aber eine semantische Nähe zu dem Cue-Item aufwiesen, wurden in der Vortest-Bedingung fälschlicherweise als bekannte, alte Items eingeschätzt. In den Experimenten 6 und 7 konnte erstmals gefunden werden, dass ein Vortest-Effekt auch dann zustande kommt, wenn statt des originalen, ursprünglichen Cue-Items (z. B. Mutter) ein semantischer Mediator als Cue-Item verwendet wird (z. B. Vater). Insgesamt deuten die Ergebnisse der Experimente 5 bis 7 also darauf hin, dass sich die positiven Effekte von Vortests auch auf inzidentelles Lernen und den Einsatz neuer, semantisch verwandter Abruf-Cues generalisieren lassen, womit eine bisher noch bestehende Lücke in der Vortest-Literatur adressiert werden konnte.

Die Experimente 5-7 wurden außerdem durchgeführt, um Erkenntnisse über die zugrundeliegenden Mechanismen des Vortest-Effekts zu sammeln und dabei insbesondere die Rolle von Elaborationsprozessen für die Entstehung dieses Effekts zu untersuchen, wodurch die Befunde dieser Experimente auch als Test der Elaborations-Hypothese des Vortest-Effekts verstanden werden können. Denn dieser Erklärungsansatz geht davon aus, dass es durch das Generieren eines Rateversuchs im initialen Vortest zu einer Aktivierung von Gedächtnisinhalten kommt, die mit dem Cue-Item verwandt sind und in einem finalen Test als semantische Mediatoren fungierend den Abruf des korrekten Target-Items erleichtern können. Falls also Vortests durch die Aktivierung der verwandten Informationen eine detailliertere und umfangreichere Gedächtnisspur erzeugen als das bloße Lernen der Items, dann sollte den Versuchspersonen in der Vortest-Bedingung diese Informationen

(semantische Mediatoren) in einem finalen Rekognitionstest mit einer größeren Wahrscheinlichkeit bekannt vorkommen. Dies sollte sich in einer größeren False Alarms Rate für Mediator-Items in der Vortest-Bedingung verglichen mit der Lern-Bedingung äußern. Genau dieses Ergebnismuster konnte in Experiment 5 gefunden werden, womit die erste Vorhersage der Elaborations-Hypothese bestätigt werden konnte. Da die Elaborations-Hypothese des Weiteren davon ausgeht, dass es durch die Bearbeitung eines Vortests zu einer Verknüpfung zwischen dem Cue-Item, dem semantischen Mediator und dem korrekten Target-Item kommt, ergibt sich daraus die zweite Vorhersage, dass die Präsentation eines semantischen Mediators als Cue-Item in einem finalen Cued-Recall-Test in der Vortest-Bedingung mit größerer Wahrscheinlichkeit zu einem korrekten Abruf des Target-Items kommt als in der Lern-Bedingung. Auch diese Vorhersage konnte durch das Ergebnismuster der Experimente 6 und 7 bestätigt werden, weshalb die Ergebnisse der vorliegenden Experimente insgesamt auf eine wichtige Rolle von Elaborationsprozessen für die Entstehung des Vortest-Effekts schließen lassen.

Mit dem Erklärungsansatz der Error-Correction-Hypothese (Carrier & Pashler, 1992) hingegen sind die vorliegenden Befunde nur schwer vereinbar. Laut dieser Hypothese sollte das Kreieren eines Fehlers während des Vortests und die anschließende Korrektur dieses Fehlers durch das unmittelbar darauffolgende Feedback (in Form des korrekten Target-Items) dazu führen, dass der Abrufpfad zwischen dem Cue-Item und dem inkorrekten Rateversuch geschwächt und infolgedessen der Abrufpfad zwischen Cue-Item und dem anschließend präsentierten Target-Item gestärkt werden sollte. Falls diese Fehlerkorrekturmechanismen auch zu einer Schwächung des Abrufpfades zu allen nicht relevanten zusätzlichen Informationen, wie den semantischen Mediatoren führen, sollte sich einerseits bei der Präsentation dieser Mediator-Items in einem Rekognitionstest eine niedrigere oder vergleichbare False Alarms Rate in der Vortest-Bedingung verglichen mit der Lern-Bedingung zeigen. Andererseits wäre zu erwarten, dass die Präsentation eines Mediator-Cues in einem finalen Cued-Recall-Test in der Vortest-Bedingung zu einer gleich guten oder schlechteren Abrufleistung für die Target-Items führen sollte, verglichen mit der Lern-Bedingung, da die Verknüpfung zwischen der zusätzlichen (nicht relevanten) Information und dem Target-Item geschwächt und somit der Abruf eines Target-Items bei Präsentation eines Mediator-Items erschwert werden sollte.

Neben der Elaborations-Hypothese und der Error-Correction-Hypothese gibt es noch weitere prominente Erklärungsansätze für den Vortest-Effekt, die zur Einordnung der vorliegenden Ergebnisse herangezogen werden können, wie beispielsweise die episodische Kontext-Hypothese oder die Aufmerksamkeits-Hypothese. Die Kontext-Hypothese (z. B. Metcalfe & Huelser, 2020; Overman et al., 2021) geht davon aus, dass während des Enkodierens von Informationen auch Details des räumlich-zeitlichen Kontexts, der während des Lernens vorhanden ist, zusammen mit dieser Information abgespeichert werden. Während Vortest-Items gemäß dieses Erklärungsansatzes also mit einem detailreicheren Kontext verknüpft sein sollten, da dieser sowohl Details aus dem Kontext während des Vortests als auch Details aus dem Kontext des anschließenden Lernens beinhalten sollte, sollten Items der Lern-Bedingung nur mit dem Lern-Kontext assoziiert sein. Der Vortest-Effekt kommt gemäß der episodischen Kontext-Hypothese folglich durch die Verfügbarkeit einer größeren Anzahl an Kontextmerkmalen in der Vortest-Bedingung verglichen mit der Lern-Bedingung zustande, die in einem finalen Test den Zugang zu den Target-Items erleichtern. Semantische Mediatoren hingegen sollten gemäß diesem Erklärungsansatz keine wichtige Rolle für die Entstehung des Vortest-Effekts spielen, weshalb die durchgeführten Experimente keine direkten Rückschlüsse für die Kontext-Hypothese ermöglichen.

Auch die Aufmerksamkeits-Hypothese (z. B. Potts & Shanks, 2014, Zawadzka & Hanczakowski, 2019) lässt sich durch die vorliegenden Befunde nur schwer überprüfen. Denn gemäß diesem Erklärungsansatz sollte das Bearbeiten eines Vortests das Enkodieren des nachfolgend präsentierten Target-Items verbessern, nicht aber die Verknüpfung zwischen dem Cue-Wort und dem Target-Wort verstärken (Potts et al., 2019). Dementsprechend sollte sich zwar eine bessere Erinnerungsleistung für die Target-Items in einem finalen Rekognitionstest zeigen, wie sie in den vorliegenden Ergebnissen gefunden werden konnte, die Erinnerungsleistung bezüglich der semantischen Mediatoren in einem finalen Rekognitions- oder Cued-Recall-Test hingegen sollte sich gemäß der Aufmerksamkeits-Hypothese nicht in Abhängigkeit davon, ob das Target-Item vorgetestet oder nur gelernt wurde, unterscheiden. Genau dieser Unterschied, mit einer größeren False Alarms Rate im Rekognitionstest und einer höheren Erinnerungsrate in einem finalen Cued-Recall-Test für die semantischen Mediatoren in der Vortest-Bedingung, konnte allerdings in den Experimenten 5-7 gefunden werden.

Das Ziel der vorliegenden drei Experimente war es einerseits, Erkenntnisse darüber zu gewinnen, ob Versuchspersonen auch dann noch von Vortests profitieren können, wenn der Lernprozess inzidentell stattfindet und im finalen Test auch neue, semantisch verwandte Cue-Items verwendet werden. Andererseits sollte durch die Untersuchung konkreter Annahmen der Elaborations-Hypothese die Rolle von Elaborationsprozessen für die Entstehung des Vortest-Effekts überprüft werden. Dabei lässt sich festhalten, dass die Aktivierung semantischer Mediatoren und die Verknüpfung und Integration des Cue-Target-Paares in dieses aktivierte semantische Netz einen wichtigen Mechanismus für den Vortest-Effekt darzustellen scheinen. Die vorliegenden Befunde stützen somit die theoretischen Annahmen der Elaborations-Hypothese. Neben Elaborationsprozessen werden aber auch Aufmerksamkeitsprozesse als wichtiger Faktor für den Vortesteffekt diskutiert und auch die Ergebnisse der Experimente 1-4 deuten auf eine mögliche Rolle von Aufmerksamkeit für den Vortest-Effekt hin. Da die Aufmerksamkeits-Hypothese die Ergebnisse der Experimente 5-7 nur teilweise erklären kann und diese Prozesse in den vorliegenden Experimenten aber auch nicht direkt untersucht werden konnten, wurden die beiden weiteren Experimente 8 und 9 durchgeführt. Mithilfe dieser Experimente sollten Erkenntnisse darüber gewonnen werden, ob aufmerksamkeitsressourcenfordernden Prozessen eine Rolle bei der Entstehung des Vortest-Effekts zukommt und wenn ja, in welcher Phase des Vortest-Trials diese speziell aufgewendet werden.

8. Der Einfluss geteilter Aufmerksamkeit auf Test-Effekte

8.1. Bisherige Befunde

Um zu untersuchen, ob und in welchem Ausmaß Aufmerksamkeitsprozesse zur erfolgreichen Bewältigung einer Aufgabe oder zur Entstehung eines Effekts beitragen, wird typischerweise das Paradigma der geteilten Aufmerksamkeit angewandt. Dabei wird einem Teil der Versuchspersonen ausschließlich die Zielaufgabe, d. h. die Aufgabe, für die der Beitrag von Aufmerksamkeitsressourcen betrachtet werden soll, präsentiert, und die Versuchspersonen können den vollen Fokus auf diese Aufgabe legen, weshalb diese Bedingung als Bedingung unter voller Aufmerksamkeit bezeichnet wird. Die Leistung dieser Gruppe in der Zielaufgabe wird verglichen mit einer Bedingung, in der parallel zur Zielaufgabe zusätzlich noch eine weitere Distraktoraufgabe (Zweitaufgabe) bearbeitet werden soll. Dadurch ist für die Bearbeitung der Zielaufgabe nur ein Teil der Aufmerksamkeitsressourcen verfügbar und diese Bedingung wird als Bedingung unter geteilter Aufmerksamkeit bezeichnet.

Als Zweitaufgabe können zwei unterschiedliche Arten zum Einsatz kommen: Zweitaufgaben, die materialspezifische Interferenz induzieren und Zweitaufgaben, die materialallgemeine Interferenz verursachen. Dabei gilt für erstere, dass die Zielaufgabe und die Zweitaufgabe ein ähnliches Material verwenden (z. B. Zahlen) und dementsprechend um Aufmerksamkeitsressourcen aus dem gleichen Repräsentationssystem konkurrieren. Materialallgemeine Interferenz hingegen wird erzeugt, wenn die Zielaufgabe und die Zweitaufgabe Material aus unterschiedlichen Kategorien (z. B. Wörter und Zahlen) verwenden und die beiden Aufgaben somit um allgemeine Aufmerksamkeitsressourcen konkurrieren (für Überblick siehe Mulligan, 2008). Materialspezifische Interferenz führt hier oft zu einer stärkeren Beeinträchtigung der Zielaufgabe, insbesondere der Gedächtnisleistung, als materialallgemeine Interferenz (Fernandes & Moscovitch, 2003). Für die Betrachtung des Einflusses von Aufmerksamkeitsprozessen auf Gedächtniseffekte

bzw. die Lernleistung wird in diesem Paradigma als Zielaufgabe typischerweise die Lernaufgabe, in der die Versuchspersonen mit dem Lerninhalt konfrontiert werden und sich diesen einprägen sollen, verwendet.

Es kann jedoch unterschieden werden, welcher Schritt des Lernprozesses unter geteilter Aufmerksamkeit betrachtet werden soll, mit unterschiedlichen Ergebnissen. So zeigt sich, dass geteilte Aufmerksamkeit während einer initialen Lernphase, in der die Lerninhalte präsentiert und anschließend von den Versuchspersonen enkodiert werden, zu einer starken Beeinträchtigung der späteren Erinnerungsleistung führen kann (z. B. Anderson et al., 1998). Findet die geteilte Aufmerksamkeit während des finalen Tests statt, in dem die gelernte Information abgerufen werden soll, scheint sich diese kaum negativ auf die Erinnerungsleistung auszuwirken (z. B. Craik et al., 1996, Naveh-Benjamin et al., 2006). Geteilte Aufmerksamkeit wirkt sich somit auf Enkodierprozesse scheinbar deutlich stärker aus als auf Abrufprozesse.

Es stellt sich jedoch die Frage, was passiert, wenn geteilte Aufmerksamkeit während einer Abrufübung stattfindet. Denn diese kann, wie zu Beginn der Arbeit beleuchtet, zum einen als Lernstrategie bezeichnet werden, da sie in sich ein Lernereignis darstellt, zum anderen soll hier aber auch Information abgerufen werden. Die Bearbeitung eines Tests könnte somit sowohl Enkodier- als auch Abrufprozesse beinhalten, auf die sich geteilte Aufmerksamkeit in unterschiedlicher Art und Weise auswirken. Um diese Fragestellung näher zu betrachten, führten Mulligan und Picklesimer (2016) eine Studie unter Verwendung des klassischen Test-Effekt Paradigmas durch. Die Versuchspersonen sollten dabei in einer initialen Lern-Phase Wortpaare lernen, die in der nachfolgenden Phase entweder erneut gelernt werden konnten oder in einem Cued-Recall-Test abgerufen werden sollten. Diese Phase fand entweder unter voller Aufmerksamkeit, d. h. es wurde ausschließlich die Lernaufgabe bearbeitet, oder unter geteilter Aufmerksamkeit statt, wobei die Versuchspersonen während des erneuten Lernens bzw. während des Bearbeitens des Tests noch eine weitere Aufgabe präsentiert bekamen. Abschließend fand ein finaler Cued-Recall-Test in allen Bedingungen statt. Die Ergebnisse des finalen Tests zeigten, dass die Erinnerungsleistung in der erneuten Lern-Bedingung unter geteilter Aufmerksamkeit deutlich schlechter ausfiel als unter voller Aufmerksamkeit, wohingegen in der Test-Bedingung dieser Unterschied zwischen voller und geteilter Aufmerksamkeit klein und nicht

signifikant war. Sowohl unter voller Aufmerksamkeit als auch unter geteilter Aufmerksamkeit ergab sich dabei ein Test-Effekt, der jedoch unter geteilter Aufmerksamkeit noch größer ausfiel. Dies deutet darauf hin, dass die vorteilhaften Effekte von Abrufübungen auch unter geteilter Aufmerksamkeit noch bestehen, bzw. vor den negativen Effekten geteilter Aufmerksamkeit zu schützen scheinen (Mulligan & Picklesimer, 2016). Diese Befunde ließen sich seither für unterschiedliche Retentionsintervalle, sowohl für materialspezifische als auch für materialallgemeine Interferenz (z. B. Buchin & Mulligan, 2017, 2019) und auch für verschiedene Testformate (Cued-Recall und freier Abruf) replizieren (Mulligan et al., 2022a, 2022b).

Während also in den genannten Studien in der erneuten Lern-Bedingung der typische negative Einfluss geteilter Aufmerksamkeit auf Enkodierprozesse auftrat, kam es in der Test-Bedingung nur zu einer geringfügigen Beeinträchtigung, was möglicherweise als Hinweis darauf gedeutet werden kann, dass die Bearbeitung eines Tests vorrangig einen Abrufprozess darstellt, der typischerweise robuster gegenüber geteilter Aufmerksamkeit ist. Allerdings findet im Paradigma des klassischen Test-Effekts diese Phase des ersten Tests bzw. des erneuten Lernens des Lernmaterials und dementsprechend die Manipulation der geteilten Aufmerksamkeit erst nach der initialen Lern-Phase statt, in der das Lernmaterial bereits einmal gelernt und enkodiert werden konnte. Daher bleibt die Frage offen, wie sich geteilte Aufmerksamkeit auf die spätere Erinnerungsleistung und auf die Größe des Erinnerungsvorteils von Abrufübungen auswirkt, wenn diese bereits während der ersten Enkodiermöglichkeit oder während eines Testes bei der erstmaligen Präsentation des Lernmaterials präsent ist, wie dies im Paradigma des Vortest-Effekts der Fall ist.

Bisher gibt es nur eine Studie, die den Vortest-Effekt unter geteilter Aufmerksamkeit untersucht hat. In Experiment 2 ihrer Studie verwendeten Mulligan und Buchin (2024) eine materialallgemeine Zweitaufgabe, die während des gesamten Lern-Trials, d. h. während der ersten Phase, in der der Vortest bearbeitet wird (Phase 1) und auch in der zweiten Phase, in der das korrektive Feedback präsentiert wird, parallel zur Lernaufgabe bearbeitet werden sollte (Phase 2). Abschließend fand ein Rekognitionstest statt, bei dem die präsentierten Items als bekannt oder neu eingestuft werden sollten. Dabei zeigten die Ergebnisse sowohl für die Bedingung unter voller Aufmerksamkeit als auch für die Bedingung unter geteilter Aufmerksamkeit einen Vortest-Effekt, der sich in seiner Größe nicht zwischen den beiden

Aufmerksamkeits-Bedingungen unterschied. Dies wurde von den Autoren als Hinweis darauf gewertet, dass der Vorteil von Vortests auf die spätere Erinnerungsleistung robust gegenüber einer Ablenkung durch eine Zweitaufgabe zu sein scheint und der Vortest-Effekt daher nicht auf aktive und kontrollierte Verarbeitungsprozesse, die Aufmerksamkeitsressourcen erfordern würden, zurückzuführen sei.

Allerdings wurde in der beschriebenen Studie ein anderes Testformat (Rekognitionstest) verwendet als in den bisherigen Studien zur Untersuchung des Einflusses geteilter Aufmerksamkeit auf den positiven Effekt von Abrufübungen. Zudem wurde in dieser Studie die geteilte Aufmerksamkeit während des gesamten Lern- bzw. Vortest-Trials manipuliert. Auf Grundlage dieser Ergebnisse lässt sich folglich keine Unterscheidung zwischen möglichen kognitiven Prozessen vornehmen, die während Phase 1, während Phase 2 oder während beider Phasen eines Lern-Trials stattfinden. Daher lassen sich die Ergebnisse von Mulligan und Buchin (2024) nur eingeschränkt in die theoretischen Erklärungsansätze des Vortest-Effekts einordnen. Denn diese Theorien schreiben den beiden Phasen eines Vortest-Trials eine unterschiedlich große Bedeutung zu und betonen in unterschiedlichem Ausmaß die Rolle von Aufmerksamkeitsprozessen innerhalb der beiden Phasen. Geteilte Aufmerksamkeit sollte sich, basierend auf den Erwartungen verschiedener prominenter Erklärungsansätzen, in den jeweiligen Phasen unterschiedlich auf den Vortest-Effekt auswirken.

8.2. Theoretische Erwartungen

Um Aussagen darüber treffen zu können, welche zugrundeliegenden Mechanismen zur Entstehung des Vortest-Effekts beitragen, erscheint es wichtig und sinnvoll, zu untersuchen, welche der beiden Phasen des Lern-Trials der typischen Vortest-Prozedur eine entscheidende Rolle für den Vortest-Effekt spielt. Denn diesbezüglich, und auch im Bezug auf die Wichtigkeit verfügbarer Aufmerksamkeitsressourcen während der beiden Phasen, kommen unterschiedliche Erklärungsansätze zu verschiedenen Vorhersagen.

Die Elaborations-Hypothese geht davon aus, dass es während eines Vortests in Phase 1 des Lern-Trials zu einer Aktivierung von Informationen kommt, die mit dem Cue-Item semantisch verwandt sind. Diese Informationen wiederum werden in Phase 2, wenn das korrektive Feedback präsentiert wird, in das bestehende Cue-Target-Wortpaar integriert und

können in einem späteren finalen Test als semantische Mediatoren den Abruf des Target-Items erleichtern (Cyr & Anderson, 2018; Huelser & Metcalfe, 2012). Dieser Erklärungsansatz sieht also sowohl in Phase 1 als auch in Phase 2 eine wichtige Voraussetzung für den Vortest-Effekt. Denn würde die Aktivierung der semantisch verwandten Information während Phase 1 ausbleiben, könnte diese in Phase 2 nicht in das Cue-Target-Paar integriert werden, und es sollten sowohl in der Lern-Bedingung als auch in der Vortest-Bedingung vergleichbare Verknüpfungen entstehen, die zu einer ähnlichen Erinnerungsleistung in einem finalen Test führen sollten. Kommt es hingegen zu einer Aktivierung der semantisch verwandten Information in Phase 1, jedoch ohne die Integration dieser in das Cue-Target-Paar in Phase 2, sollte diese bei einem späteren Abruf nicht als zusätzlicher Hinweisreiz verfügbar sein und der Abruf der Target-Information dadurch in der Vortest-Bedingung nicht besser ausfallen als in der Lern-Bedingung.

Ähnlich wie die Elaborations-Hypothese nimmt auch die Search-Set-Hypothese an, dass es durch den Vortest in Phase 1 zu einer Aktivierung von Gedächtnisrepräsentationen in einem Search-Set kommt, das auch die korrekte Target-Information enthält. Da diese Target-Information in Phase 1 bereits voraktiviert wurde, sollte diese in Phase 2, wenn das korrektive Feedback zu sehen ist, effizienter enkodiert werden. Dadurch sollte in einem späteren finalen Test der Abruf der Target-Items in der Vortest-Bedingung besser ausfallen als in einer Lern-Bedingung, in der diese tiefere Enkodierung ausbleibt. Gemäß dieser Hypothese könnte also eine Störung in Phase 1, durch die die Aktivierung des Search-Sets behindert wird oder ausbleibt, dazu führen, dass die Target-Information in der Vortest-Bedingung in Phase 2 nicht effizienter enkodiert wird als dies in der Lern-Bedingung der Fall ist, wodurch sich in einem späteren finalen Test auch die Abrufleistung der beiden Bedingungen nicht unterscheiden sollte. Wird trotz einer Voraktivierung des Search-Sets in Phase 1 die Enkodierung des Target-Items in Phase 2 gestört, könnte auch dies dazu führen, dass der Erinnerungsvorteil in der Vortest-Bedingung, der durch diese tiefere Enkodierung der Target-Information entsteht, ausbleibt. Folglich könnten auch gemäß der Search-Set-Hypothese beide Phasen eine entscheidende Rolle für den Vortest-Effekt spielen. Bei beiden zuvor beschriebenen Erklärungsansätze ist jedoch unklar, ob und inwiefern die in beiden Phasen ablaufenden Prozesse Aufmerksamkeitsressourcen erfordern. Daher lassen sich aus

diesen Erklärungsansätzen keine konkreten Erwartungen bezüglich des Einflusses geteilter Aufmerksamkeit in den beiden Phasen auf die Größe des Vortest-Effekts ableiten.

Gemäß der Kontext-Hypothese (Metcalf & Huelser, 2020) basiert der Vortest-Effekt auf einer größeren Verfügbarkeit hilfreicher Kontextmerkmale für Vortest-Items als für Items der Lern-Bedingung, die den Abruf der korrekten Target-Information erleichtern. Während Items der Lern-Bedingung nur mit Merkmalen des Lern-Kontexts verknüpft werden, sind Items in der Vortest-Bedingung sowohl mit dem Vortest-Kontext als auch mit dem Lern-Kontext assoziiert. In der Vortest-Bedingung stehen im finalen Test somit mehr Kontextmerkmale zur Verfügung, die durch eine Reaktivierung den Zugriff auf die Zielinformation fördern und so zu einer besseren Erinnerungsleistung in der Vortest-Bedingung beitragen. Da es sich bei der Integration der räumlich-zeitlichen Kontextmerkmale während des Enkodierens des Lernmaterials um einen überwiegend automatisierten und unbewussten Prozess handelt, sollte sich gemäß der Kontext-Hypothese geteilte Aufmerksamkeit durch eine Zweitaufgabe weder in Phase 1 noch in Phase 2 direkt auf die spätere Erinnerungsleistung auswirken. Somit lässt sich der Beitrag der Kontext-Hypothese für den Vortest-Effekt durch die Experimente 8 und 9 nicht konkret überprüfen.

Ein anderer prominenter Erklärungsansatz hingegen lässt eine differenzierte Erwartung bezüglich der Rolle von Aufmerksamkeitsressourcen und des Einflusses geteilter Aufmerksamkeit zu, da dieser vor allem in einer der beiden Phasen die Ursache für den Vortest-Effekt sieht.

So geht die Aufmerksamkeits-Hypothese des Vortest-Effekts davon aus, dass insbesondere Phase 2 des Lern-Trials vorrangig für die Entstehung des Effekts entscheidend ist (Potts & Shanks, 2014; Zawadzka & Hanczakowski, 2019). Dieser Erklärungsansatz nimmt an, dass Versuchspersonen durch die Bearbeitung des Vortests und das Generieren eines eigenen Rateversuchs darauf aufmerksam gemacht werden, dass ein Informationsdefizit zwischen der gesuchten Target-Information und dem eigenen Wissensstand besteht. Das Vorliegen dieser Diskrepanz soll einerseits die Neugierde der Versuchspersonen für das korrekte Target-Item erhöhen und andererseits zu einer gesteigerten Motivation führen, die korrekte Lösung zu lernen, um die aufgedeckte Wissenslücke zu füllen. Dies führt gemäß der Aufmerksamkeits-Hypothese dazu, dass Versuchspersonen mehr Aufmerksamkeitsressourcen darauf verwenden, das Target-Item zu

enkodieren, wenn dieses während Phase 2 des Lern-Trials in Form von korrektivem Feedback bereitgestellt wird. Diese intensivere und aufmerksamere Verarbeitung des Target-Items in der Vortest-Bedingung wiederum trägt dazu bei, dass in einem späteren finalen Test das gesuchte Target-Item besser korrekt abgerufen werden kann als in der Lern-Bedingung, in der diese stärkere Enkodierung nicht stattfindet, wodurch der Vortest-Effekt zustande kommt.

Die Aufmerksamkeits-Hypothese geht folglich davon aus, dass insbesondere in Phase 2 des Lern-Trials aufmerksamkeitsressourcenfordernde Prozesse bei der Enkodierung des Feedbacks verantwortlich für die Entstehung des Effekts sind, wohingegen das Feststellen des Informationsdefizites während Phase 1 einen eher automatisierten Prozess darstellt, der nur geringfügig Aufmerksamkeit erfordert. Es sollte also beim Bearbeiten eines Vortests folglich auch dann noch zu der Feststellung einer Diskrepanz zwischen dem eigenen Wissensstand und der gesuchten Zielinformation und dementsprechend zu einer größeren Neugierde und Motivation für die Verarbeitung der Target-Information kommen, wenn während dieser Phase die Aufmerksamkeit durch eine Zweitaufgabe geteilt wird. In der Vortest-Bedingung sollte sich also geteilte Aufmerksamkeit in Phase 1 kaum oder nur geringfügig negativ auf die spätere Erinnerungsleistung auswirken, wohingegen in der Lern-Bedingung die Bearbeitung einer Zweitaufgabe während des Enkodierens zu dem typischen negativen Einfluss geteilter Aufmerksamkeit führen sollte.

In Summe sollte folglich der Vortest-Effekt gemäß der Aufmerksamkeits-Hypothese durch das Teilen der Aufmerksamkeit in Phase 1 gleich bleiben oder sogar noch größer werden als unter voller Aufmerksamkeit. Findet hingegen die geteilte Aufmerksamkeit während Phase 2 des Lern-Trials statt, sollte dies sich in der Lern-Bedingung weniger stark auswirken als in Phase 1, da hier die Zielinformation bereits für eine gewisse Zeit unter voller Aufmerksamkeit enkodiert werden konnte. In der Vortest-Bedingung sollte eine geteilte Aufmerksamkeit in Phase 2 aber die Fokussierung von Aufmerksamkeitsressourcen auf die Enkodierung der Target-Information, wenn diese in Phase 2 präsentiert wird, reduzieren oder gar verhindern, wodurch die spätere Erinnerungsleistung an die Target-Information beeinträchtigt werden sollte. Insgesamt würde die Aufmerksamkeits-Hypothese für eine geteilte Aufmerksamkeit in Phase 2 folglich einen geringeren oder gar ausbleibenden Vortest-Effekt erwarten.

Auch die Error-Correction-Hypothese sieht den Ursprung des Vortest-Effekts vorrangig in kognitiven Prozessen, die in Phase 2 des Lern-Trials ablaufen. In dieser Phase führt die Präsentation des korrektiven Feedbacks zur Entstehung eines Fehlersignals, das überwiegend automatisiert und unbewusst ausgelöst wird (z. B. Hester et al., 2005). Ausgehend von diesem Fehlersignal werden Fehlerkorrekturmechanismen aktiviert, die die fehlerhafte Abrufroute zwischen dem Cue-Item und dem Rateversuch inhibieren und zugleich die Abrufroute zum korrekten Target-Item stärken. Diese beiden Vorgänge scheinen jedoch kognitive Kontrolle und Aufmerksamkeit zu erfordern (z. B. Botvinick et al., 2001). Eine Ablenkung in Phase 2 durch eine Zweitaufgabe könnte folglich zu einer Reduktion der verfügbaren Aufmerksamkeitsressourcen führen und bedingen, dass die fehlerhafte Abrufroute nicht (ausreichend) gehemmt wird. Zusätzlich sollte es dadurch auch zu einer ausbleibenden Stärkung der Verknüpfung zwischen dem Cue-Target-Paar kommen, wodurch der positive Effekt von Vortests abgeschwächt wird. Gemäß der Error-Correction-Hypothese wäre es folglich zu erwarten, dass eine geteilte Aufmerksamkeit in Phase 2 des Lern-Trials zu einer Reduktion des Vortest-Effekts führt, wohingegen sich geteilte Aufmerksamkeit in Phase 1 nicht auf die Größe des Vortest-Effekts auswirken sollte.

8.3. Fragestellung der Experimente 8-9

Bisherige Studien zum Einfluss geteilter Aufmerksamkeit auf späteres Erinnern legen die mögliche Schlussfolgerung nahe, dass sich ein Aufteilen der Aufmerksamkeitsressourcen unterschiedlich auf die spätere Merkfähigkeit für das Lernmaterial auswirkt, je nachdem, ob dabei Enkodierprozesse oder Abrufprozesse durch eine Zweitaufgabe gestört werden. Insbesondere beim Paradigma des Vortest-Effekts finden jedoch sowohl Abrufprozesse in Form der Bearbeitung eines Vortests in der Phase 1 des Lern-Trials als auch Enkodierprozesse durch die Verarbeitung des in Phase 2 des Lern-Trials präsentierten korrektiven Feedbacks innerhalb der Lern-Phase statt. Wie sich jedoch geteilte Aufmerksamkeit auf dieses Paradigma auswirkt, wurde bisher nur in der Studie von Mulligan und Buchin (2024) untersucht. In dieser Studie fand, wie zuvor beschrieben, die geteilte Aufmerksamkeit jedoch über die komplette Dauer des Lern-Trials statt. Aufgrund dessen ermöglicht diese Vorgängerstudie keine direkten Rückschlüsse darüber, welchen Einfluss

geteilte Aufmerksamkeit in jeder der beiden Phasen separat auf die Größe des Vortest-Effekts hat.

Daher war es das Ziel der nachfolgenden Experimente 8 und 9, zu untersuchen, inwiefern sich geteilte Aufmerksamkeit während des initialen Vortests (Phase 1) und während der anschließenden Feedback-Phase (Phase 2) auf die Größe des Vortest-Effekts auswirkt. Dies würde erste Rückschlüsse darauf ermöglichen, in welcher der beiden Phasen möglicherweise aufmerksamkeitsressourcenfördernde Prozesse eine zentrale Rolle für die Entstehung des Vortest-Effekts spielen und in welcher der beiden Phasen somit gezielt auf eine Vermeidung von Ablenkung geachtet werden sollte. Dazu wurden den Versuchspersonen in beiden Experimenten die zu lernenden Wortpaare entweder vollständig präsentiert, wodurch diese direkt gelernt werden konnten (Lern-Bedingung) oder es wurde zunächst nur das Cue-Wort präsentiert und die Versuchspersonen sollten versuchen, das dazugehörige Target-Wort zu erraten, bevor anschließend das vollständige Wortpaar zu sehen war (Vortest-Bedingung). Die Hälfte der Lern-Trials fand dabei unter voller Aufmerksamkeit statt, wohingegen die andere Hälfte der Lern-Trials unter geteilter Aufmerksamkeit, d. h. in Anwesenheit einer Zweitaufgabe, stattfanden. Die Zweitaufgabe wurde parallel zur Lern-Aufgabe präsentiert, und bestand dabei entweder in der Klassifikation von Zahlen als gerade oder ungerade Zahlen (Experiment 8), was materialallgemeine Interferenz induzieren sollte, oder in der Zuteilung von Nomen auf die beiden Kategorien „vom Menschen gemacht“ oder „von der Natur geschaffen“ (Experiment 2). Dies sollte materialspezifische Interferenz produzieren und damit eine noch größere Konkurrenz für die Lern-Aufgabe darstellen. Die Zweitaufgabe wurde dabei entweder während Phase 1 eines Lern-Trials (Experimente 8a und 9a) oder während Phase 2 des Lern-Trials (Experiment 8b und 9b) präsentiert. In einem späteren finalen Test wurde die Erinnerungsleistung der Versuchspersonen an alle Wortpaare dadurch erfasst, dass den Versuchspersonen jeweils die Cue-Wörter aller Wortpaare vorgegeben wurden und das dazugehörige Target-Wort genannt werden sollte.

In beiden Experimenten wurde dabei in der Bedingung unter voller Aufmerksamkeit ein klassischer Vortest-Effekt mit einer besseren Erinnerungsleistung für vorgetestete Wortpaare verglichen mit nur gelernten Wortpaaren erwartet, was eine Replikation vorheriger Studien darstellen würde (z. B. Kornell et al., 2009). Weniger klar war die

Vorhersage bezüglich des Einflusses geteilter Aufmerksamkeit in den Phasen 1 und 2 auf die Größe des Vortest-Effekts. Aus einigen Erklärungsansätzen, wie der Search-Set-Hypothese oder der Elaborations-Hypothese, die beiden Phasen eines Lern-Trials eine wichtige Rolle für die Entstehung des Vortest-Effekts beimessen, lassen sich keine konkreten Erwartungen ableiten, wie sich geteilte Aufmerksamkeit in den beiden unterschiedlichen Phasen eines Lern-Trials auf den Vortest-Effekt auswirken könnte. Ebenso ergeben sich auch aus der Kontext-Hypothese keine direkt überprüfbaren Annahmen über die Rolle geteilter Aufmerksamkeit. Falls der Vortest-Effekt allerdings vorrangig durch aufmerksamkeitsressourcenfordernde Prozesse in Phase 2 des Lern-Trials zustande kommen sollte, wie dies die Aufmerksamkeits-Hypothese oder auch die Error-Correction-Hypothese des Vortest-Effekts erwarten ließen, sollte das Bearbeiten einer Zweitaufgabe in Phase 1 kaum Auswirkungen auf den Vortest-Effekt haben, wohingegen eine Ablenkung durch eine Zweitaufgabe in Phase 2 zu einer Reduktion des Vortest-Effekts führen sollte. Die Ergebnisse der Experimente 8 und 9 stellen somit einerseits vor allem einen direkten Test der Aufmerksamkeits-Hypothese des Vortest-Effekts dar, andererseits tragen sie zu einem besseren Verständnis für die Rolle aufmerksamkeitsressourcenfordernder kognitiver Prozesse für den Vortest-Effekt bei.

9. Der Vortest-Effekt unter geteilter Aufmerksamkeit

9.1. Experimente 8a und 8b

Method

Da die Experimente 8a und 8b abgesehen von kleinen prozeduralen Unterschieden identisch waren, werden diese im Folgenden gemeinsam beschrieben.

Ethische Berücksichtigungen. Die nachfolgend berichteten Experimente 8 und 9 wurden in Übereinstimmung mit den Vorgaben der Deklaration von Helsinki des Weltärztebundes durchgeführt.

Stichprobe. Um die notwendige Stichprobengröße der Experimente 8 und 9 zu bestimmen, wurde eine Poweranalyse mittels G*Power (Version 3.1.9.2; Faul et al., 2009) durchgeführt. Als Parameter wurden hierbei ein α von .05, ein β von .20 und eine Effektgröße von Hedges' g von 0.44 basierend auf der metaanalytischen Schätzung der Größe des Vortest-Effekts (Boustani & Shanks, 2022) verwendet, wobei sich eine notwendige Stichprobengröße von $N = 43$ ergab. Die tatsächliche Stichprobe lag mit 52 Versuchspersonen sowohl für Experiment 8a ($M = 25.3$ Jahre, Altersspanne = 18-39 Jahre, 35 weiblich) als auch für Experiment 8b ($M = 25.1$ Jahre, Altersspanne = 19-36 Jahre, 39 weiblich) etwas darüber. Alle Versuchspersonen sprachen Deutsch als Muttersprache und gaben ihr mündliches Einverständnis zur Teilnahme. Alle Versuchspersonen erhielten als Aufwandsentschädigung entweder Versuchspersonenstunden oder eine monetäre Vergütung.

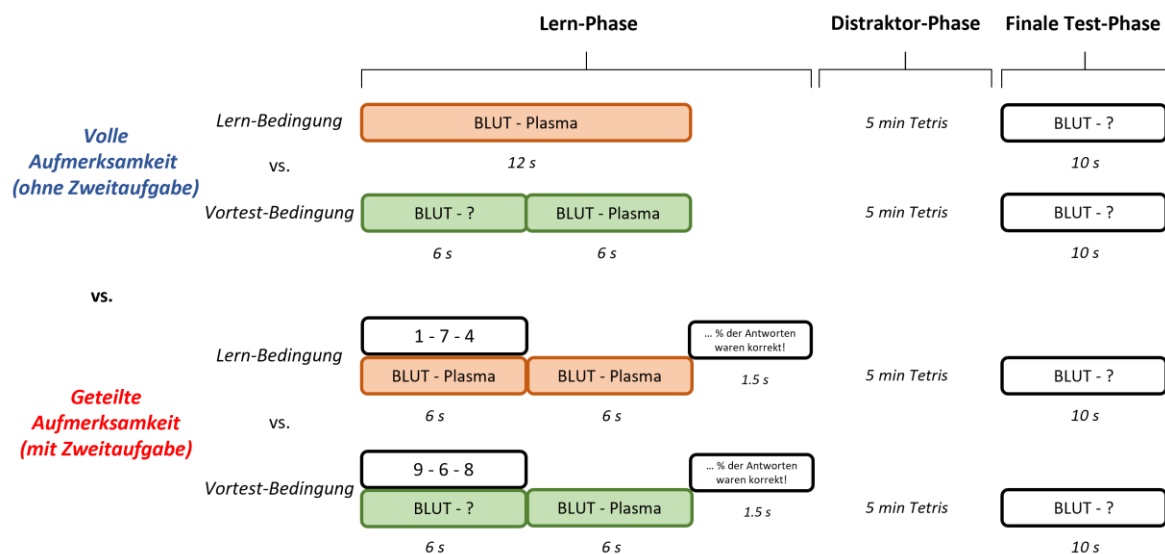
Material. Als Lernmaterial wurden 60 schwach miteinander verwandte Wortpaare (z. B. Blut - Plasma, Zug - Flugzeug) verwendet, die aus den Normen von Nelson et al. (1998) entnommen wurden. Die vorwärtsgerichtete Assoziationsstärke betrug dabei im Mittel einen Wert von .052 und variierte zwischen .047 und .067. Das bedeutet, dass in der Ursprungsstudie von Nelson et al. die Versuchspersonen in etwa 5 % der Fälle das korrekte

Target-Wort als erste Antwort gaben, wenn ihnen ein Cue-Wort präsentiert wurde. Alle Items wurden ins Deutsche übersetzt. Für jede Versuchsperson wurden die 60 Wortpaare zufällig in 2 Untergruppen mit jeweils 30 Wortpaaren unterteilt und jede Untergruppe wurde einem der beiden Experimentalblöcke zugeordnet. Innerhalb der beiden Blöcke wurden 15 der Wortpaare zufällig in Vortest-Trials und die anderen 15 Wortpaare in Lern-Trials verwendet.

Design und Versuchsablauf. Die Experimente 8a und 8b wurden online mittels der Videotelefonie-Software Zoom (Zoom Video Communications Inc., 2016) durchgeführt, wobei alle Versuchspersonen in Einzeltestungen, d. h. in separaten Videokonferenzen, getestet wurden. Beide Experimente bestanden aus drei unterschiedlichen Phasen: Einer Lern-Phase, einer Distraktor-Phase und einer finalen Test-Phase. In Abbildung 6 ist der Ablauf von Experiment 8a beispielhaft dargestellt.

Abbildung 6

Versuchsablauf von Experiment 8a



Anmerkungen: Der Versuchsablauf unterteilt in Lern-Phase, Distraktor-Phase und finale Test-Phase. Ein Block der Lern-Phase wurde dabei von allen Versuchspersonen unter voller Aufmerksamkeit absolviert, der andere Block fand unter geteilter Aufmerksamkeit statt. Im finalen Test wurden alle gelernten Wortpaare in einem Cued-Recall-Test abgefragt.

Die Lern-Phase bestand aus zwei Blöcken mit jeweils 30 Lern-Trials, wobei jeder Lern-Trial wiederum in zwei Phasen unterteilt war: Phase 1, welche die ersten 6 s des Lern-Trials beinhaltete und Phase 2, die die zweiten 6 s eines Lern-Trials umfasste. Die Hälfte der Cue-Wörter eines jeden Blocks wurden in Phase 1 zusammen mit dem entsprechenden Target-Wort präsentiert (z. B. Blut – *Plasma*) und konnten direkt gelernt werden (Lern-Bedingung), wohingegen für die andere Hälfte der Wortpaare in Phase 1 zunächst für 6 s nur das Cue-Wort präsentiert wurde (z. B. Zug - ?) und die Versuchspersonen versuchen sollten, das dazugehörige Target-Wort zu erraten (Vortest-Bedingung). Während Phase 2, die ohne Unterbrechung auf Phase 1 folgte, wurde in beiden Bedingungen für 6 s das intakte Wortpaar gezeigt (z. B. Blut – *Plasma*; Zug – *Flugzeug*). Während in Trials der Lern-Bedingung die Phasen 1 und 2 lediglich einen einzelnen insgesamt 12 s umfassenden Abschnitt darstellten, in dem das vollständige Wortpaar von Beginn an gelernt werden konnte, erfüllte Phase 2 in der Vortest-Bedingung die Funktion eines 6 s andauernden Feedbacks, das auf den Vortest, der in Phase 1 stattfand, folgte. Alle Wortpaare dieser Lern-Aufgabe wurden dabei für jede Versuchsperson in einer zufälligen Reihenfolge präsentiert und in der unteren Hälfte des Bildschirms angezeigt.

Für jede Versuchsperson fand einer der beiden Experimentalblöcke (wie oben beschrieben) unter voller Aufmerksamkeit statt, während der andere Block unter geteilter Aufmerksamkeit absolviert wurde. Die Hälfte der Versuchspersonen durchlief dabei den ersten Block unter voller Aufmerksamkeit und den folgenden zweiten Block unter geteilter Aufmerksamkeit, wohingegen diese Reihenfolge der Blöcke für die andere Hälfte der Versuchspersonen genau umgekehrt war. Der Block mit geteilter Aufmerksamkeit war identisch zum Block mit voller Aufmerksamkeit mit dem einzigen Unterschied, dass die Versuchspersonen aufgefordert wurden, zusätzlich noch eine Zweitaufgabe zu bearbeiten. Diese Zweitaufgabe fand entweder während Phase 1 (Experiment 8a) oder Phase 2 (Experiment 8b) des Lern-Trials statt und sollte von den Versuchspersonen parallel zur Lern-Aufgabe (Erstaufgabe) bearbeitet werden. Zu Beginn des Blocks mit geteilter Aufmerksamkeit wurden die Versuchspersonen darüber informiert, dass die Zahlen-Kategorisierungs-Aufgabe (Zweitaufgabe) genauso wichtig war wie die Lern-Aufgabe (Erstaufgabe). Für die Zweitaufgabe wurden in jedem Trial zufällig drei Zahlen (1 - 9, ohne 5) der Reihe nach für jeweils 2 s in der oberen Bildschirmhälfte präsentiert. Die Aufgabe der

Versuchspersonen bestand darin, die Zahlen so schnell und korrekt wie möglich als gerade Zahlen, durch das Drücken der „g“-Taste auf der Tastatur, oder als ungerade Zahlen, durch das Drücken der „u“-Taste auf der Tastatur, zu kategorisieren. Am Ende eines jeden Trials erhielten die Versuchspersonen für 1.5 s ein Feedback über die Genauigkeit bei der Bearbeitung der Zweitaufgabe (z. B. „67 % der Antworten waren korrekt“, falls zwei von drei Antworten korrekt waren), wobei verpasste Antworten, d. h., wenn innerhalb der 2 s dauernden Präsentation einer Zahl kein Tastendruck abgegeben wurde, als Fehler gewertet wurden. Im Anschluss an das Feedback folgte der nächste Trial.

Nachdem beide Blöcke der Lern-Phase absolviert wurden, folgte eine 5-minütige Distraktor-Phase, in der alle Versuchspersonen eine Online-Version des Videospiels Tetris (<https://www.geo.de/geolino/spiele/13349-rtklonlinespiel-tetris>) spielen sollten, bevor der finale Test stattfand. In diesem finalen Cued-Recall-Test wurden die Cue-Wörter aller 60 Wortpaare für 10 s in einer zufälligen Reihenfolge präsentiert und die Versuchspersonen sollten das dazugehörige Target-Wort abrufen (z. B. Blut - ?). Es erfolgte im finalen Test kein Feedback über die Korrektheit der Antwort. Alle Antworten im Vortest sowie im finalen Test wurden dabei mündlich gegeben und vom Versuchsleiter notiert. Für die Experimente 8a und 8b ergab sich folglich ein 2 x 2 Design mit den Innersubjekt-Faktoren ÜBUNG (Lern-Bedingung vs. Vortest-Bedingung) und AUFMERKSAMKEIT (volle Aufmerksamkeit vs. geteilte Aufmerksamkeit).

Ergebnisse von Experiment 8a

Initialer Vortest. Im initialen Vortest zeigte sich kein signifikanter Unterschied im Anteil an korrekt erratenen Target-Wörtern zwischen Vortest-Trials unter voller Aufmerksamkeit und Vortest-Trials unter geteilter Aufmerksamkeit (4.9 % vs. 4.4 %), $t(51) < 1$. Da das Hauptinteresse dem Effekt falscher Rateversuche auf die nachfolgende Gedächtnisleistung galt, wurden alle Wortpaare, für die im Vortest das korrekte Target-Wort erraten wurde, von den nachfolgenden Analysen ausgeschlossen.

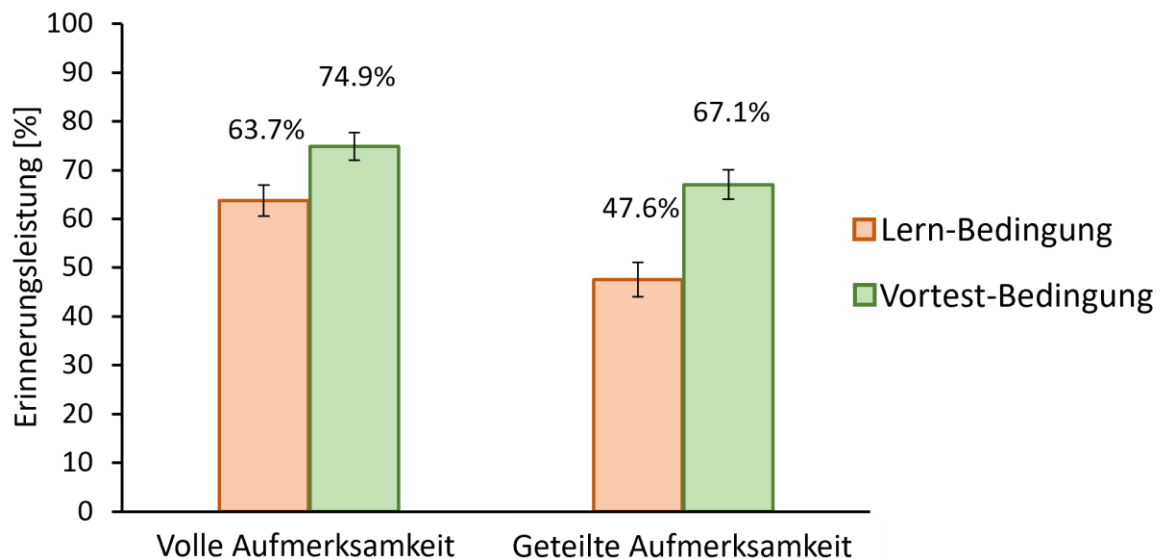
Finaler Test.

Korrechter Abruf. Abbildung 7 zeigt den prozentualen Anteil an korrekt erinnerten Target-Items im finalen Test in Abhängigkeit der Faktoren ÜBUNG (Lern-Bedingung vs. Vortest-Bedingung) und AUFMERKSAMKEIT (volle Aufmerksamkeit vs. geteilte

Aufmerksamkeit). Eine 2 x 2 ANOVA der beiden Faktoren ergab einen signifikanten Haupteffekt ÜBUNG, $F(1,51) = 62.79$, $MSE = 133.76$, $p < .001$, $\eta^2_p = .55$, mit einer besseren Erinnerungsleistung für Target-Items der Vortest-Bedingung verglichen mit der Lern-Bedingung (70.9 % vs. 55.6 %) und einen signifikanten Haupteffekt AUFMERKSAMKEIT, $F(1,51) = 23.85$, $MSE = 133.76$, $p < .001$, $\eta^2_p = .32$, mit einem größeren Anteil korrekt erinnelter Items für Wortpaare, die unter voller Aufmerksamkeit gelernt wurden, verglichen mit Wortpaaren, die unter geteilter Aufmerksamkeit präsentiert wurden (69.3 % vs. 57.3 %). Zusätzlich konnte auch eine signifikante Interaktion der beiden Faktoren gefunden werden, $F(1,51) = 6.76$, $MSE = 133.76$, $p = .012$, $\eta^2_p = .12$. Geplante paarweise Vergleiche ergaben eine bessere Erinnerungsleistung für Target-Items in der Vortest-Bedingung verglichen mit der Lern-Bedingung sowohl für Wortpaare, die unter voller Aufmerksamkeit gelernt wurden (74.9 % vs. 63.7 %), $t(51) = 5.60$, $p < .001$, Cohen's $d = 0.78$, als auch für Wortpaare, die unter geteilter Aufmerksamkeit gelernt wurden (67.1 % vs. 47.6 %), $t(51) = 6.62$, $p < .001$, Cohen's $d = 0.92$. Jedoch nahm die Größe des Vortest-Effekts von 11.1 % unter voller Aufmerksamkeit auf 19.5 % unter geteilter Aufmerksamkeit zu.

Abbildung 7

Ergebnisse von Experiment 8a



Anmerkung: Erinnerungsleistung im finalen Test (in %) in Abhängigkeit der beiden Faktoren Bedingung (Lernen vs. Vortest) und Aufmerksamkeit (volle Aufmerksamkeit vs. geteilte Aufmerksamkeit). Fehlerbalken repräsentieren den Standardfehler des Mittelwertes.

Beschränkte Analysen. In der Bedingung mit geteilter Aufmerksamkeit bearbeiteten die Versuchspersonen die Zahlen-Kategorisierungs-Aufgabe (Zweitaufgabe) während parallel dazu die Lern-Aufgabe (Erstaufgabe) absolviert werden sollte. Ein paarweiser Vergleich zeigte eine signifikant bessere Leistung in der Zweitaufgabe für die Lern-Bedingung verglichen mit der Vortest-Bedingung (89.9 % vs. 78.5 %), $t(51) = 6.31, p < .001$, Cohen's $d = 0.88$. Um auszuschließen, dass die oben berichtete Interaktion zwischen den beiden Faktoren ÜBUNG und AUFMERKSAMKEIT nur dadurch zustande kam, dass in der Vortest-Bedingung mehr Aufmerksamkeitsressourcen in die Lern-Aufgabe investiert wurden als in der Lern-Bedingung, wurde eine beschränkte Analyse durchgeführt. In diese Analyse wurden nur die Wortpaare aus Trials aufgenommen, in denen die Versuchspersonen in der Zweitaufgabe eine Korrektheit von 100 % erzielten (für ein ähnliches Vorgehen siehe Buchin & Mulligan, 2017)². Unter Verwendung dieses reduzierten Daten-Sets zusammen mit den Daten aus den Trials unter voller Aufmerksamkeit (diese blieben identisch, da hier keine Zweitaufgabe bearbeitet wurde), wurde erneut die oben berichtete 2 x 2 ANOVA durchgeführt. Dabei ergab sich wie für die Analyse der vollständigen Daten eine signifikante Interaktion der beiden Faktoren, $F(1,50) = 5.21, MSE = 245.50, p = .027, \eta^2_p = .09$. Diese Interaktion legt nahe, dass sich auch dann noch ein größerer Vortest-Effekt unter geteilter Aufmerksamkeit als unter voller Aufmerksamkeit zeigt, wenn die Leistung der Zweitaufgabe in der Lern-Bedingung und der Vortest-Bedingung gleich war.

Intrusionen. Erneut wurden alle fehlerhaften Antworten, die im finalen Test von den Versuchspersonen gegeben wurden, als Intrusionen gewertet. Eine 2 x 2 ANOVA mit den beiden Faktoren ÜBUNG (Lern-Bedingung vs. Vortest-Bedingung) und AUFMERKSAMKEIT (volle Aufmerksamkeit vs. geteilte Aufmerksamkeit) bezüglich der Anzahl an Intrusionen ergab signifikante Haupteffekte ÜBUNG, $F(1,51) = 33.31, MSE = 2.33, p < .001, \eta^2_p = .40$, und AUFMERKSAMKEIT, $F(1,51) = 12.24, MSE = 2.33, p < .001, \eta^2_p = .19$, mit einer insgesamt größeren Anzahl an Intrusionen in der Lern-Bedingung verglichen mit der Vortest-Bedingung (4.4 Intrusionen vs. 2.7 Intrusionen) und weniger produzierten Intrusionen unter voller Aufmerksamkeit als unter geteilter Aufmerksamkeit

² Versuchspersonen, die in keinem der Trials der Vortest- oder Lern-Bedingung eine Leistung von 100 % in der Zweitaufgabe erreichen konnten, wurden aus dem Daten-Set ausgeschlossen. Das betraf in den Experimenten 8a, 8b und 9a jeweils eine Versuchsperson, in Experiment 9b insgesamt vier Versuchspersonen.

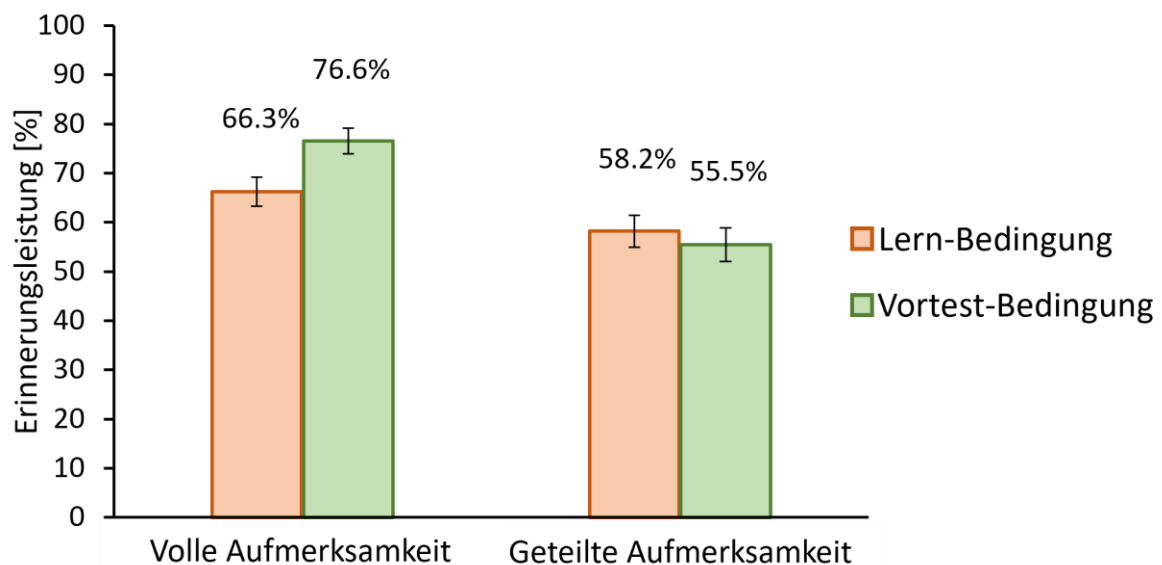
(3.1 Intrusionen vs. 4.0 Intrusionen). Es zeigte sich zudem auch eine signifikante Interaktion der beiden Faktoren, $F(1,51) = 6.47$, $MSE = 2.33$, $p = .014$, $\eta^2_p = .11$. Paarweise Vergleiche offenbarten, dass es in der Vortest-Bedingung sowohl unter voller Aufmerksamkeit, $t(51) = 3.68$, $p < .001$, Cohen's $d = 0.51$, als auch unter geteilter Aufmerksamkeit, $t(51) = 5.44$, $p < .001$, Cohen's $d = 0.76$, zu weniger Intrusionen kam als in der Lern-Bedingung, und dass die Größe dieses Vorteils von Vortests von 1.2 Intrusionen (2.5 Intrusionen vs. 3.7 Intrusionen) unter voller Aufmerksamkeit auf 2.2 Intrusionen (2.9 Intrusionen vs. 5.1 Intrusionen) unter geteilter Aufmerksamkeit anstieg.

Ergebnisse von Experiment 8b

Initialer Vortest. Bezüglich der Leistung im initialen Vortest ergab sich für die Vortest-Bedingung erneut kein signifikanter Unterschied zwischen Trials unter voller Aufmerksamkeit und Trials unter geteilter Aufmerksamkeit (5.1 % vs. 4.1 %), $t(51) = 1.16$, $p = .25$, Cohen's $d = 0.16$. Alle Wortpaare, für die im Vortest das korrekte Target-Wort erraten wurde, wurden von den anschließenden Analysen ausgeschlossen.

Finaler Test.

Korrektur Abruf. In Abbildung 8 ist der prozentuale Anteil korrekt erinnelter Target-Wörter im finalen Test in Abhängigkeit der Faktoren ÜBUNG (Lern-Bedingung vs. Vortest-Bedingung) und AUFMERKSAMKEIT (volle Aufmerksamkeit vs. geteilte Aufmerksamkeit) zu sehen. Eine 2 x 2 ANOVA der beiden Faktoren zeigte einen signifikanten Haupteffekt AUFMERKSAMKEIT, $F(1,51) = 36.66$, $MSE = 122.17$, $p < .001$, $\eta^2_p = .42$, wobei Wortpaare, die unter voller Aufmerksamkeit gelernt wurden, besser erinnert werden konnten als Wortpaare, die unter geteilter Aufmerksamkeit gelernt wurden (71.4 % vs. 56.9 %), jedoch keinen Haupteffekt ÜBUNG, $F(1,51) = 3.40$, $MSE = 122.17$, $p = .071$, $\eta^2_p = .06$. Wichtig war vor allem aber der Befund einer signifikanten Interaktion der beiden Faktoren, $F(1,51) = 18.02$, $MSE = 122.17$, $p < .001$, $\eta^2_p = .26$. Dabei zeigten geplante paarweise Vergleiche unter voller Aufmerksamkeit eine bessere Erinnerungsleistung in der Vortest-Bedingung als in der Lern-Bedingung (76.6 % vs. 66.3 %), $t(51) = 4.45$, $p < .001$, Cohen's $d = 0.62$, was einen Vortest-Effekt von 10.3 % widerspiegelt. Unter geteilter Aufmerksamkeit hingegen unterschieden sich die Vortest-Bedingung und die Lern-Bedingung nicht signifikant (55.5 % vs. 58.2 %), $t(51) < 1$.

Abbildung 8*Ergebnisse von Experiment 8b*

Anmerkung: Erinnerungsleistung im finalen Test (in %) in Abhängigkeit der beiden Faktoren Bedingung (Lernen vs. Vortest) und Aufmerksamkeit (volle Aufmerksamkeit vs. geteilte Aufmerksamkeit). Fehlerbalken repräsentieren den Standardfehler des Mittelwertes.

Beschränkte Analysen. Wie in Experiment 8a, war auch in Experiment 8b die Leistung in der Zweitaufgabe in der Lern-Bedingung signifikant besser als in der Vortest-Bedingung (87.7 % vs. 76.0 %), $t(51) = 7.35$, $p < .001$, Cohen's $d = 1.02$, weshalb erneut eine beschränkte Analyse durchgeführt wurde, in die für die Bedingung mit geteilter Aufmerksamkeit nur Trials mit einer 100 % korrekten Antwort in der Zweitaufgabe einbezogen wurden. Diese 2 x 2 ANOVA mit den beschränkten Daten ergab genau wie die Analyse der vollständigen Daten eine signifikante Interaktion zwischen den beiden Faktoren, $F(1,50) = 8.97$, $MSE = 174.39$, $p = .004$, $\eta^2_p = .15$, was einen signifikanten Vortest-Effekt unter voller Aufmerksamkeit, nicht aber unter geteilter Aufmerksamkeit anzeigt.

Intrusionen. Zur Analyse der Anzahl an Intrusionen wurde eine 2 x 2 ANOVA mit den Faktoren ÜBUNG (Lern-Bedingung vs. Vortest-Bedingung) und AUFMERKSAMKEIT (volle Aufmerksamkeit vs. geteilte Aufmerksamkeit) durchgeführt, wobei sich signifikante Haupteffekte ÜBUNG, $F(1,51) = 10.61$, $MSE = 1.94$, $p = .002$, $\eta^2_p =$

.17, und AUFMERKSAMKEIT, $F(1,51) = 27.24$, $MSE = 1.94$, $p < .00$, $\eta^2_p = .35$, ergaben. Insgesamt führten Vortest-Trials zu einer geringeren Anzahl an Intrusionen als Lern-Trials (2.9 Intrusionen vs. 3.6 Intrusionen), und in Trials unter voller Aufmerksamkeit wurden weniger Intrusionen produziert als in Trials unter geteilter Aufmerksamkeit (2.6 Intrusionen vs. 3.9 Intrusionen). Es konnte zudem eine signifikante Interaktion der beiden Faktoren gefunden werden, $F(1,51) = 7.51$, $MSE = 1.94$, $p = .008$, $\eta^2_p = .13$. Paarweise Vergleiche offenbarten dabei, dass unter voller Aufmerksamkeit in der Vortest-Bedingung weniger Intrusionen erzeugt wurden als in der Lern-Bedingung (2.0 Intrusionen vs. 3.3 Intrusionen), $t(51) = 4.49$, $p < .001$, Cohen's $d = 0.62$. Dieser Vorteil von Vortests war aber unter geteilter Aufmerksamkeit nicht vorhanden (3.8 Intrusionen vs. 4.0 Intrusionen), $t(51) < 1$.

Zusätzliche Analysen

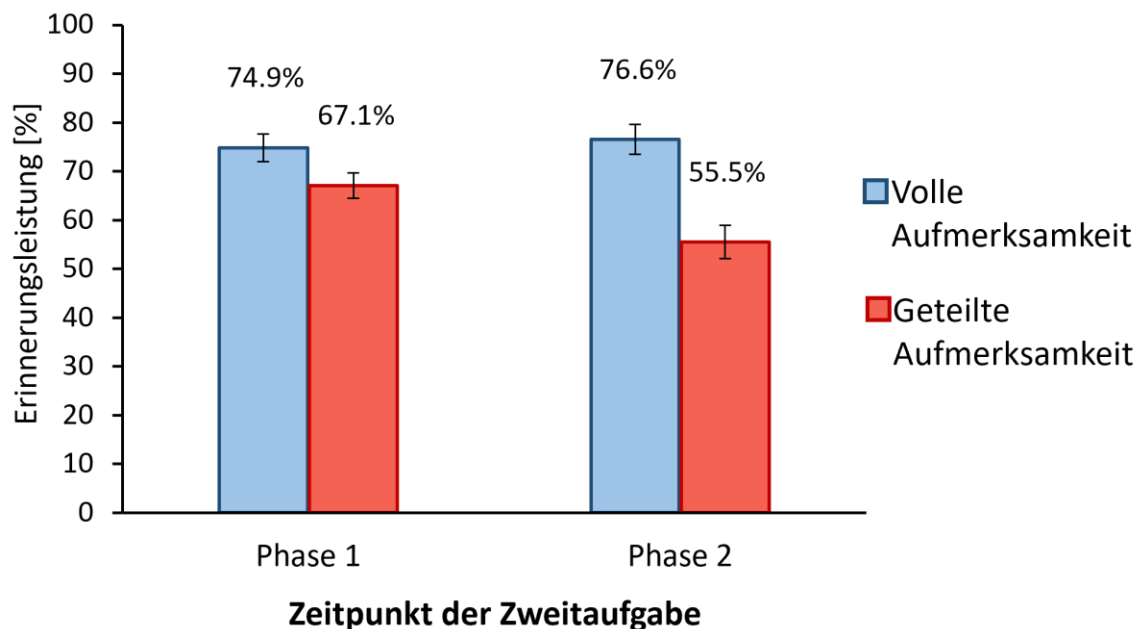
Bei Betrachtung der Erinnerungsleistung der Lern-Bedingung im finalen Test zeigte sich, dass diese um 10.6 % niedriger ausfiel, wenn die Zweitaufgabe in Phase 1 des Lern-Trials stattfand (Experiment 8a), als wenn die Zweitaufgabe in Phase 2 des Lern-Trials zu bearbeiten war (Experiment 8b). Daher wäre es möglich, dass das beobachtete Ergebnismuster aus den Experimenten 8a (größerer Vortest-Effekt unter geteilter Aufmerksamkeit) und 8b (ausbleibender Vortest-Effekt unter geteilter Aufmerksamkeit) zumindest teilweise auf Unterschiede in der Baseline Bedingung (Lern-Bedingung) zurückzuführen war. Um diese Möglichkeit auszuschließen, wurde eine zusätzliche Analyse durchgeführt, die lediglich die zusammengefassten Daten der Vortest-Bedingung der Experimente 8a und 8b, nicht aber die Daten der Lern-Bedingung beinhaltete.

Eine 2 x 2 ANOVA mit den Faktoren AUFMERKSAMKEIT (volle Aufmerksamkeit vs. geteilte Aufmerksamkeit) und ZEITPUNKT DER ZWEITAUFGABE (Phase 1 vs. Phase 2) zeigte einen signifikanten Haupteffekt AUFMERKSAMKEIT, $F(1,102) = 51.98$, $MSE = 208.99$, $p < .001$, $\eta^2_p = .34$, mit einer besseren Erinnerungsleistung in der Bedingung unter voller Aufmerksamkeit als in der Bedingung unter geteilter Aufmerksamkeit (75.7 % vs. 61.3 %), aber keinen Haupteffekt ZEITPUNKT DER ZWEITAUFGABE, $F(1,102) = 1.77$, $MSE = 714.71$, $p = .19$, $\eta^2_p = .02$. Vor allem aber konnte eine signifikante Interaktion der beiden Faktoren gefunden werden, $F(1,102) = 10.96$, $MSE = 208.99$, $p < .001$, $\eta^2_p = .10$. Die Erinnerungsleistung für die Bedingung unter geteilter Aufmerksamkeit war verglichen mit der Bedingung unter voller Aufmerksamkeit sowohl dann beeinträchtigt, wenn die

Zweitaufgabe während Phase 1 stattfand als auch, wenn die Zweitaufgabe in Phase 2 präsentiert wurde (alle $ps < .004$). Dieser Unterschied zwischen der Bedingung unter voller Aufmerksamkeit und der Bedingung unter geteilter Aufmerksamkeit fiel dann größer aus, wenn die Zweitaufgabe in Phase 2 statt in Phase 1 stattfand (21.1 % vs. 7.8 %) (siehe Abbildung 9). Diese Ergebnisse legen in Kombination mit den Befunden der zuvor berichteten Analysen nahe, dass das Bearbeiten einer Zweitaufgabe in Phase 2 zu einer stärkeren Beeinträchtigung der Erinnerungsleistung in der Vortest-Bedingung führte, als dies bei der Präsentation der Zweitaufgabe in Phase 1 der Fall war.

Abbildung 9

Ergebnisse der Zusatzanalyse der Experimente 8a und 8b



Anmerkung: Erinnerungsleistung (in %) im finalen Test in Abhängigkeit der beiden Faktoren Zeitpunkt der Zweitaufgabe (Phase 1 vs. Phase 2) und Aufmerksamkeit (volle Aufmerksamkeit vs. geteilte Aufmerksamkeit). Fehlerbalken repräsentieren den Standardfehler des Mittelwertes.

Diskussion

In beiden Experimenten 8a und 8b konnte unter voller Aufmerksamkeit ein Erinnerungsvorteil für Material, für das ein Vortest bearbeitet werden sollte, verglichen mit

Information, die ausschließlich gelernt wurde, gefunden werden, was den typischen Vortest-Effekt darstellt. Noch wichtiger war allerdings der Befund, dass das Bearbeiten einer Zweitaufgabe parallel zur Lern-Aufgabe zu einem Verschwinden des Vortest-Effekts führen kann, wenn diese Zweitaufgabe während Phase 2 eines Lern-Trials stattfindet (Experiment 8b), nicht aber, wenn die Zweitaufgabe in Phase 1 präsentiert wird (Experiment 8a). Im Gegenteil zeigten die Ergebnisse von Experiment 8a, dass die Bearbeitung einer Zweitaufgabe in Phase 1 sogar zu einer Vergrößerung des Vortest-Effekts führte. Dies kam dadurch zustande, dass die Erinnerungsleistung in der Lern-Bedingung durch die geteilte Aufmerksamkeit stärker beeinträchtigt war als die Erinnerungsleistung in der Vortest-Bedingung. Für beide Experimente 8a und 8b zeigte sich das gleiche Ergebnismuster, unabhängig davon, ob alle Items in die Analysen miteinbezogen wurden oder ob nur Items, für die die Zweitaufgabe vollständig korrekt beantwortet wurde, berücksichtigt wurden. So offenbarte die beschränkte Analyse in Experiment 8a, dass die Zunahme des Vortest-Effekts nicht einfach dadurch zustande kam, dass die Versuchspersonen in der Vortest-Bedingung mehr Aufmerksamkeitsressourcen auf die Lern-Aufgabe (Erstaufgabe) verwendeten als in der Lern-Bedingung. Denn wäre dies der Fall gewesen, hätte sich die Zunahme des Vortest-Effekts für Items, bei denen von einer annähernd gleichen Verteilung der Aufmerksamkeitsressourcen auf die Zweit- und somit auch auf die Erst-(Lern-)Aufgabe für die Vortest-Bedingung und die Lern-Bedingung ausgegangen werden kann, nicht mehr zeigen dürfen.

Des Weiteren ergab die Analyse der Intrusions-Daten eine größere Anzahl an Intrusionen im finalen Test in der Lern-Bedingung als in der Vortest-Bedingung, wenn die Lern-Trials unter voller Aufmerksamkeit stattfanden, was die Ergebnisse vorheriger Studien, die ebenfalls diesen vorteilhaften Effekt von Vortests zeigen konnten, repliziert (z. B. Kliegl et al., 2024). Fanden die Lern-Trials unter geteilter Aufmerksamkeit statt, führte das Bearbeiten von Vortests nur dann zu einer reduzierten Anzahl an Intrusionen, wenn die Zweitaufgabe während Phase 1, nicht aber wenn die Zweitaufgabe während Phase 2 präsentiert wurde, was komplementär zu den Befunden der Erinnerungsleistung war. Beide vorteilhaften Effekte von Vortests ergaben sich demnach in der Bedingung mit geteilter Aufmerksamkeit nur, wenn die Zweitaufgabe in Phase 1 präsentiert wurde. Zuletzt ließ sich mittels einer zusätzlichen Analyse, die ausschließlich die zusammengefassten Daten der

Vortest-Bedingungen der beiden Experimente betrachtete, zeigen, dass die Erinnerungsleistung unter geteilter Aufmerksamkeit verglichen mit der Bedingung unter voller Aufmerksamkeit sowohl, wenn die Zweitaufgabe in Phase 1, als auch, wenn die Zweitaufgabe in Phase 2 stattfand, beeinträchtigt war. Diese Beeinträchtigung war jedoch deutlich stärker, wenn die Versuchspersonen in Phase 2 durch eine weitere Aufgabe abgelenkt wurden, als wenn diese Zweitaufgabe in Phase 1 präsentiert wurde. Dies legt die Vermutung nahe, dass es im Besonderen Phase 2 zu sein scheint, die Aufmerksamkeitsressourcen erfordert, die für das Entstehen des Vortest-Effekts eine wichtige Rolle spielen.

Das Ziel von Experiment 9 war es, die Ergebnisse von Experiment 8 konzeptuell zu replizieren. Dabei unterschieden sich die beiden Experimente nur in der Art der Zweitaufgabe, die in Experiment 8 aus einer Zahlen-Klassifikations-Aufgabe bestand, wohingegen in Experiment 9 eine Wort-Klassifikations-Aufgabe verwendet wurde. Diese Aufgabe wurde verwendet, um zu untersuchen, ob sich die Ergebnisse von Experiment 8 auch auf die Induktion von materialspezifischer Interferenz generalisieren lassen, bei der die Zweitaufgabe und die Lern-Aufgabe (Erstaufgabe) um Aufmerksamkeitsressourcen aus dem gleichen Repräsentationssystem konkurrieren. Da materialspezifische Interferenz eine noch größere Ablenkung darstellen sollte als die materialallgemeine Interferenz, ist es unklar, ob geteilte Aufmerksamkeit unter diesen Bedingungen zu dem gleichen Ergebnismuster führt wie in Experiment 8.

9.2. Experimente 9a und 9b

Method

Wie bereits in Experiment 8 unterschieden sich auch die Experimente 9a und 9b nur in kleinen prozeduralen Details. Dementsprechend werden die beiden Experimente im Anschluss gemeinsam beschrieben.

Stichprobe. Der Poweranalyse aus Experiment 8 folgend nahmen jeweils 52 Versuchspersonen an Experiment 9a ($M = 25.3$ Jahre, Altersspanne = 18-37, 35 weiblich) und an Experiment 9b ($M = 24.5$ Jahre, Altersspanne = 18-35, 35 weiblich) teil. Für alle Versuchspersonen war Deutsch die Muttersprache und alle Versuchspersonen gaben ihr

mündliches Einverständnis. Als Gegenleistung für die Teilnahme erhielten die Versuchspersonen entweder Versuchspersonenstunden oder eine monetäre Vergütung.

Material, Design und Versuchsablauf. Die Experimente 9a und 9b waren identisch zu den Experimenten 8a und 8b mit dem einzigen Unterschied, dass statt einer Zahlen-Klassifikations-Aufgabe eine Wort-Klassifikations-Aufgabe als Zweitaufgabe verwendet wurde. Dafür wurden 90 Nomen verwendet, von denen die Hälfte als „vom Menschen gemacht“ (z. B. Uhr, Ofen, Kanu) und die andere Hälfte als „von der Natur geschaffen“ (z. B. Tomate, Schmetterling, Teich) klassifiziert werden konnten. Wie in Experiment 8 fand diese Zweitaufgabe entweder während Phase 1 des Lern-Trials (Experiment 9a) oder während Phase 2 des Lern-Trials (Experiment 9b) statt. Während eines jeden Trials wurden dabei drei zufällige Nomen der Reihe nach in der oberen Bildschirmhälfte präsentiert, wobei jedes Nomen für 2s sichtbar war. Die Aufgabe der Versuchspersonen bestand darin, die Nomen so schnell und korrekt wie möglich durch das Drücken der „m“-Taste auf der Tastatur als „vom Menschen gemacht“ und durch das Drücken der „n“-Taste auf der Tastatur als „von der Natur geschaffen“ zu klassifizieren. Insgesamt wurden dabei Nomen der beiden Kategorien gleich häufig an jeder der möglichen Stellen während der Zweitaufgabe präsentiert.

Ergebnisse von Experiment 9a

Initialer Vortest. Während des initialen Vortests unterschied sich der Anteil an korrekt erratenen Target-Items nicht signifikant zwischen der Bedingung unter voller Aufmerksamkeit und der Bedingung unter geteilter Aufmerksamkeit (2.8 % vs. 4.4 %), $t(51) = 1.63$, $p = .11$, Cohen's $d = 0.23$. Alle Wortpaare, für die das korrekte Target-Item erraten wurde, wurden von den nachfolgenden Analysen ausgeschlossen, da der Einfluss falscher Rateversuche auf die spätere Erinnerung von zentralem Interesse war.

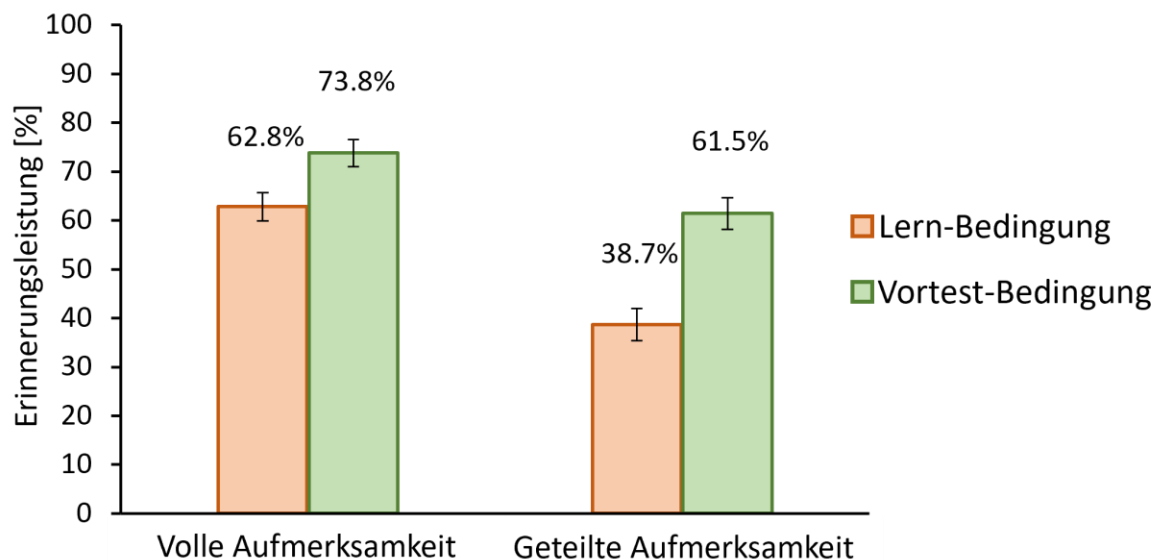
Finaler Test.

Korrektur Abruf. Abbildung 10 zeigt den prozentualen Anteil an korrekt erinnerten Target-Items im finalen Test in Abhängigkeit der Faktoren ÜBUNG (Lern-Bedingung vs. Vortest-Bedingung) und AUFMERKSAMKEIT (volle Aufmerksamkeit vs. geteilte Aufmerksamkeit). Eine 2 x 2 ANOVA der beiden Faktoren zeigte signifikante Haupteffekte

ÜBUNG, $F(1,51) = 52.84$, $MSE = 211.29$, $p < .001$, $\eta^2_p = .51$, und AUFMERKSAMKEIT, $F(1,51) = 63.47$, $MSE = 211.29$, $p < .001$, $\eta^2_p = .55$, mit einer insgesamt besseren Erinnerungsleistung für Target-Items der Vortest-Bedingung verglichen mit den Target-Items der Lern-Bedingung (67.6 % vs. 50.8 %) und mit einem insgesamt größeren Anteil korrekt erinnelter Items für Wortpaare, die unter voller Aufmerksamkeit gelernt wurden, verglichen mit Wortpaaren, die unter geteilter Aufmerksamkeit präsentiert wurden (68.3 % vs. 50.1 %). Vor allem ergab sich aber auch eine signifikante Interaktion der beiden Faktoren, $F(1,51) = 8.50$, $MSE = 211.29$, $p = .005$, $\eta^2_p = .14$. Geplante paarweise Vergleiche ergaben eine bessere Erinnerungsleistung in der Vortest-Bedingung verglichen mit der Lern-Bedingung sowohl für Wortpaare, die unter voller Aufmerksamkeit gelernt wurden (73.8 % vs. 62.8 %), $t(51) = 4.03$, $p < .001$, Cohen's $d = 0.56$, als auch für Wortpaare, die unter geteilter Aufmerksamkeit gelernt wurden (61.5 % vs. 38.7 %), $t(51) = 6.72$, $p < .001$, Cohen's $d = 0.93$. Insgesamt stieg folglich die Größe des Vortest-Effekts von 11.0 % unter voller Aufmerksamkeit auf 22.7 % unter geteilter Aufmerksamkeit an.

Abbildung 10

Ergebnisse von Experiment 9a



Anmerkung: Erinnerungsleistung im finalen Test (in %) in Abhängigkeit der beiden Faktoren Bedingung (Lernen vs. Vortest) und Aufmerksamkeit (volle Aufmerksamkeit vs. geteilte Aufmerksamkeit). Fehlerbalken repräsentieren den Standardfehler des Mittelwertes.

Beschränkte Analysen. Wie in den zuvor berichteten Experimenten 8a und 8b zeigte ein paarweiser Vergleich eine bessere Leistung in der Zweitaufgabe in der Lern-Bedingung als in der Vortest-Bedingung (81.8 % vs. 65.0 %), $t(51) = 10.99$, $p < .001$, Cohen's $d = 1.52$. Daher wurde erneut eine beschränkte Analyse durchgeführt, die lediglich Items enthielt, für die in der Zweitaufgabe eine Korrektheit von 100 % erzielt wurde. Eine 2 x 2 ANOVA der beschränkten Daten ergab wie die Analyse der vollständigen Daten eine signifikante Interaktion der beiden Faktoren $F(1,50) = 5.52$, $MSE = 430.08$, $p = .023$, $\eta^2_p = .10$, womit sich auch dann ein größerer Vortest-Effekt für Trials unter geteilter Aufmerksamkeit als für Trials unter voller Aufmerksamkeit zeigte, wenn die Leistung in der Zweitaufgabe für beide Bedingungen gleich gut war.

Intrusionen. Zur Analyse der Anzahl an Intrusionen wurde eine 2 x 2 ANOVA mit den Faktoren ÜBUNG (Lern-Bedingung vs. Vortest-Bedingung) und AUFMERKSAMKEIT (volle Aufmerksamkeit vs. geteilte Aufmerksamkeit) durchgeführt, wobei signifikante Haupteffekte ÜBUNG, $F(1,51) = 50.24$, $MSE = 2.46$, $p < .001$, $\eta^2_p = .50$, und AUFMERKSAMKEIT, $F(1,51) = 62.42$, $MSE = 2.46$, $p < .001$, $\eta^2_p = .55$, gefunden werden konnten. Diese Haupteffekte zeigten an, dass weniger Intrusionen in der Vortest-Bedingung als in der Lern-Bedingung gemacht wurden (3.0 Intrusionen vs. 5.0 Intrusionen), und dass Trials unter voller Aufmerksamkeit zu weniger Intrusionen führten als Trials unter geteilter Aufmerksamkeit (2.9 Intrusionen vs. 5.1 Intrusionen). Zudem ergab sich eine signifikante Interaktion der beiden Faktoren, $F(1,51) = 6.34$, $MSE = 2.46$, $p = .015$, $\eta^2_p = .11$. Paarweise Vergleiche zeigten dabei, dass in der Vortest-Bedingung weniger Intrusionen produziert wurden als in der Lern-Bedingung, unabhängig davon, ob die Trials unter voller Aufmerksamkeit (2.2 Intrusionen vs. 3.6 Intrusionen), $t(51) = 4.43$, $p < .001$, Cohen's $d = 0.61$, oder unter geteilter Aufmerksamkeit stattfanden (3.8 Intrusionen vs. 6.3 Intrusionen), $t(51) = 6.54$, $p < .001$, Cohen's $d = 0.91$. Jedoch stieg auch die Größe dieses Vorteils von Vortests von 1.4 Intrusionen unter voller Aufmerksamkeit auf 2.5 Intrusionen unter geteilter Aufmerksamkeit an.

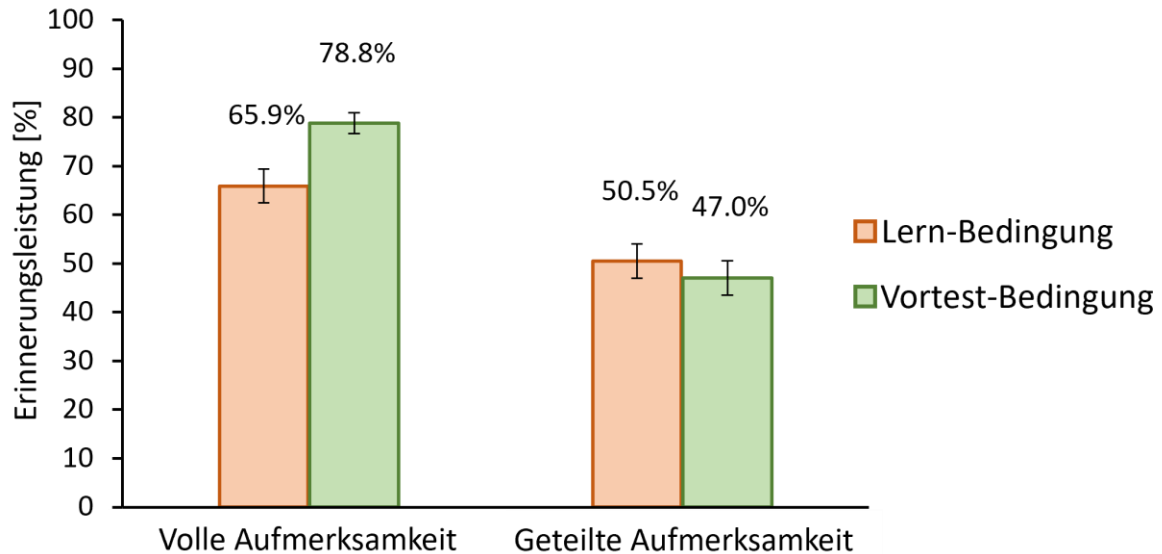
Ergebnisse von Experiment 9b

Initialer Vortest. Im initialen Vortest ergab sich kein signifikanter Unterschied im Anteil der korrekt erratenen Target-Wörter zwischen den Vortest-Trials unter voller

Aufmerksamkeit und den Vortest-Trials unter geteilter Aufmerksamkeit (5.6 % vs. 5.9 %), $t(51) < 1$. Erneut wurden alle Wortpaare, für die im initialen Vortest die korrekte Antwort erraten wurde, von den nachfolgenden Analysen ausgeschlossen.

Abbildung 11

Ergebnisse von Experiment 9b



Anmerkung: Erinnerungsleistung im finalen Test (in %) in Abhängigkeit der beiden Faktoren Bedingung (Lernen vs. Vortest) und Aufmerksamkeit (volle Aufmerksamkeit vs. geteilte Aufmerksamkeit). Fehlerbalken repräsentieren den Standardfehler des Mittelwertes.

Finaler Test.

Korrektur Abruf. Der prozentuale Anteil an korrekt erinnerten Target-Wörtern im finalen Test in Abhängigkeit der Faktoren ÜBUNG (Lern-Bedingung vs. Vortest-Bedingung) und AUFMERKSAMKEIT (volle Aufmerksamkeit vs. geteilte Aufmerksamkeit) ist in Abbildung 11 dargestellt. Eine 2 x 2 ANOVA der beiden Faktoren zeigte einen signifikanten Haupteffekt ÜBUNG, $F(1,51) = 4.42$, $MSE = 147.35$, $p = .041$, $\eta^2_p = .08$, mit einer insgesamt besseren Erinnerungsleistung in der Vortest-Bedingung verglichen mit der Lern-Bedingung (62.9 % vs. 58.2 %) und einen signifikanten Haupteffekt AUFMERKSAMKEIT, $F(1,51) = 97.70$, $MSE = 147.35$, $p < .001$, $\eta^2_p = .66$, mit einem größeren Anteil korrekt erinnerten Target-Wörter für Wortpaare, die unter voller

Aufmerksamkeit gelernt wurden verglichen mit Wortpaaren, die unter geteilter Aufmerksamkeit gelernt wurden (72.4 % vs. 48.8 %). Zudem zeigte sich auch eine signifikante Interaktion der beiden Faktoren, $F(1,51) = 23.79$, $MSE = 147.35$, $p < .001$, $\eta^2_p = .32$. Paarweise Vergleiche offenbarten dabei, dass in der Vortest-Bedingung mehr Target-Wörter korrekt erinnert wurden als in der Lern-Bedingung, wenn die Lern-Trials unter voller Aufmerksamkeit stattfanden (78.8 % vs. 65.9 %), $t(51) = 4.51$, $p < .001$, Cohen's $d = 0.63$, was einen klassischen Vortest-Effekt von 12.9 % widerspiegelte. Fanden die Lern-Trials hingegen unter geteilter Aufmerksamkeit statt, verschwand dieser vorteilhafte Effekt von Vortests, und es zeigte sich kein signifikanter Unterschied zwischen der Vortest- und der Lern-Bedingung mehr (47.0 % vs. 50.5 %), $t(51) = 1.28$, $p = .21$, Cohen's $d = 0.18$.

Beschränkte Analysen. Bei der Betrachtung der Leistung in der Zweitaufgabe offenbarte sich wie in den bisherigen Experimenten eine bessere Leistung in der Lern-Bedingung verglichen mit der Vortest-Bedingung (84.4 % vs. 71.2 %), $t(51) = 7.38$, $p < .001$, Cohen's $d = 1.02$. Aus diesem Grund, wurde auch für Experiment 9b eine beschränkte Analyse durchgeführt, in die nur die Wortpaare aus der Bedingung mit geteilter Aufmerksamkeit einbezogen wurden, für welche die Zweitaufgabe mit 100 % Korrektheit beantwortet wurde. Diese beschränkte 2 x 2 ANOVA zeigte ebenso wie die Analyse der vollständigen Daten eine signifikante Interaktion der beiden Faktoren, $F(1,47) = 7.13$, $MSE = 205.97$, $p = .01$, $\eta^2_p = .13$, was ebenfalls einen Vortest-Effekt für die Bedingung unter voller Aufmerksamkeit, nicht aber für die Bedingung mit geteilter Aufmerksamkeit, ergab.

Intrusionen. Um die Anzahl an Intrusionen in Abhängigkeit der Faktoren ÜBUNG (Lern-Bedingung vs. Vortest-Bedingung) und AUFMERKSAMKEIT (volle Aufmerksamkeit vs. geteilte Aufmerksamkeit) zu untersuchen, wurde eine 2 x 2 ANOVA der beiden Faktoren durchgeführt. Dabei ergab sich ein signifikanter Haupteffekt ÜBUNG, $F(1,51) = 7.66$, $MSE = 2.81$, $p = .008$, $\eta^2_p = .13$, wobei in der Vortest-Bedingung weniger Intrusionen produziert wurden als in der Lern-Bedingung (3.6 Intrusionen vs. 4.3 Intrusionen) und ein signifikanter Haupteffekt AUFMERKSAMKEIT, $F(1,51) = 63.55$, $MSE = 2.81$, $p < .001$, $\eta^2_p = .56$, mit einer höheren Anzahl an Intrusionen in Trials unter geteilter Aufmerksamkeit als in Trials unter voller Aufmerksamkeit (5.1 Intrusionen vs. 2.8 Intrusionen). Es zeigte sich zudem eine signifikante Interaktion der beiden Faktoren, $F(1,51) = 10.41$, $MSE = 2.81$, $p = .002$, $\eta^2_p = .17$. Paarweise Vergleiche offenbarten dabei, dass in

Trials unter voller Aufmerksamkeit in der Vortest-Bedingung signifikant weniger Intrusionen produziert wurden als in der Lern-Bedingung (2.0 Intrusionen vs. 3.5 Intrusionen), $t(51) = 3.94$, $p < .001$, Cohen's $d = 0.55$. Dieser Vorteil von Vortests verschwand jedoch, wenn die Trials beider Bedingungen unter geteilter Aufmerksamkeit absolviert wurden (5.1 Intrusionen vs. 5.0 Intrusionen), $t(51) < 1$.

Zusätzliche Analysen

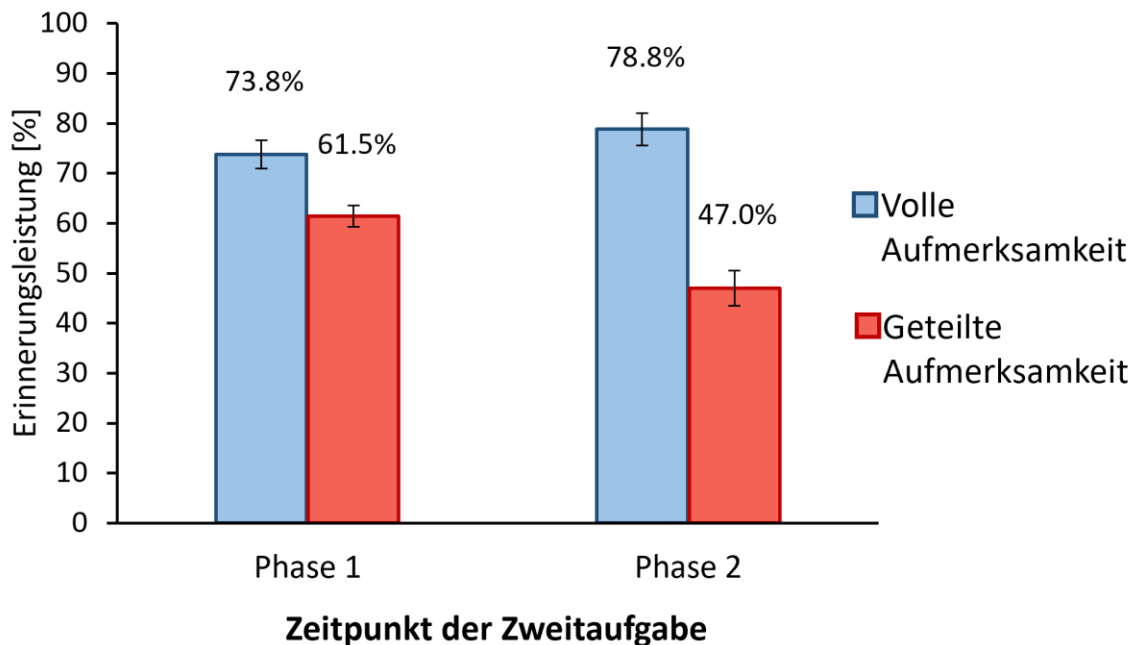
In der Lern-Bedingung fiel auch in Experiment 9, ähnlich wie in Experiment 8, der Anteil korrekt erinnelter Target-Wörter im finalen Test um 11.8 % geringer aus, wenn die Zweitaufgabe in Phase 1 (Experiment 9a) des Lern-Trials stattfand, als wenn die Zweitaufgabe in Phase 2 des Lern-Trials präsentiert wurde (Experiment 9b). Um die Möglichkeit auszuschließen, dass diese Unterschiede in der Erinnerungsleistung der Lern-Bedingung zu dem beobachteten Ergebnismuster der Experimente 9a und 9b beigetragen haben, wurde erneut eine zusätzliche ANOVA durchgeführt, die lediglich die zusammengefassten Daten der Vortest-Bedingung aus den Experimenten 9a und 9b berücksichtigte.

Die 2×2 ANOVA mit den Faktoren AUFMERKSAMKEIT (volle Aufmerksamkeit vs. geteilte Aufmerksamkeit) und ZEITPUNKT DER ZWEITAUFGABE (Phase 1 vs. Phase 2) zeigte einen signifikanten Haupteffekt AUFMERKSAMKEIT, $F(1,102) = 89.19$, $MSE = 284.14$, $p < .001$, $\eta^2_p = .47$, mit einer insgesamt besseren Erinnerungsleistung in der Bedingung unter voller Aufmerksamkeit als in der Bedingung unter geteilter Aufmerksamkeit (76.3 % vs. 54.2 %), aber keinen signifikanten Haupteffekt ZEITPUNKT DER ZWEITAUFGABE, $F(1,102) = 1.82$, $MSE = 637.11$, $p = .18$, $\eta^2_p = .02$. Vor allem aber ergab sich eine signifikante Interaktion der beiden Faktoren, $F(1,102) = 17.32$, $MSE = 284.14$, $p < .001$, $\eta^2_p = .15$. Sowohl, wenn die Zweitaufgabe in Phase 1, als auch, wenn die Zweitaufgabe in Phase 2 des Lern-Trials präsentiert wurde, war die Erinnerungsleistung in der Bedingung mit geteilter Aufmerksamkeit verglichen mit der Bedingung mit voller Aufmerksamkeit beeinträchtigt (alle $ps < .001$). Jedoch war diese Beeinträchtigung in der Bedingung mit geteilter Aufmerksamkeit gegenüber der Bedingung mit voller Aufmerksamkeit deutlich größer, wenn die Zweitaufgabe in Phase 2 statt in Phase 1 stattfand (31.8 % vs. 12.3 %) (siehe Abbildung 12). Diese Ergebnisse passen folglich gut zu den

Befunden der zuvor berichteten Analysen, die auch die Daten der Lern-Bedingung beinhalteten.

Abbildung 12

Ergebnisse der Zusatzanalyse der Experimente 9a und 9b



Anmerkung: Erinnerungsleistung (in %) im finalen Test in Abhängigkeit der beiden Faktoren Zeitpunkt der Zweitaufgabe (Phase 1 vs. Phase 2) und Aufmerksamkeit (volle Aufmerksamkeit vs. geteilte Aufmerksamkeit). Fehlerbalken repräsentieren den Standardfehler des Mittelwertes.

Diskussion

Wie vorab erwartet, führte die Verwendung einer materialspezifischen Zweitaufgabe in Experiment 9 zu einem größeren Ausmaß an Interferenz als die materialallgemeine Zweitaufgabe in Experiment 8. Verglichen mit der Bedingung unter voller Aufmerksamkeit war die Erinnerungsleistung an die Target-Wörter im finalen Test in der Lern-Bedingung in Experiment 9 (materialspezifische Interferenz) insgesamt stärker beeinträchtigt als in Experiment 8 (materialallgemeine Interferenz) (19.7 % vs. 12.1 %). Abgesehen davon, stellten die Ergebnisse von Experiment 9 eine Replikation der Befundmuster aus Experiment 8 dar. So zeigten die Ergebnisse von Experiment 9 einen typischen Vortest-Effekt unter

voller Aufmerksamkeit, der unter geteilter Aufmerksamkeit entweder in seiner Größe noch zunahm, wenn die Zweitaufgabe in Phase 1 des Lern-Trials stattfand (Experiment 9a) oder durch die Anwesenheit einer Zweitaufgabe in Phase 2 des Lern-Trials (Experiment 9b) eliminiert wurde. Wie bereits in Experiment 8 konnte dieses Ergebnismuster nicht nur bei der Betrachtung der vollständigen Daten gefunden werden, sondern auch, wenn in der Analyse nur Wortpaare berücksichtigt wurden, für die die Zweitaufgabe vollständig korrekt beantwortet wurde. Dies legt nahe, dass der gefundene Anstieg in der Größe des Vortest-Effekts in Experiment 9a nicht nur auf eine strategische Verteilung der Aufmerksamkeitsressourcen in der Vortest-Bedingung zurückzuführen ist.

Zusätzlich konnte in der Bedingung mit geteilter Aufmerksamkeit einmal mehr nur dann eine geringere Anzahl an Intrusionen im finalen Test in der Vortest-Bedingung verglichen mit der Lern-Bedingung gefunden werden, wenn die Zweitaufgabe während Phase 1, nicht aber, wenn die Zweitaufgabe während Phase 2 eines Lern-Trials, präsentiert wurde. Zuletzt wurde auch für Experiment 9 eine zusätzliche Analyse durchgeführt, in die nur die zusammengefassten Daten der Vortest-Bedingung aus den Experimenten 9a und 9b eingeschlossen wurden. Die Ergebnisse dieser Analyse ergaben, wie in Experiment 8, dass die Erinnerung an die korrekten Target-Wörter im finalen Test im Vergleich zu der jeweiligen Bedingung unter voller Aufmerksamkeit stärker beeinträchtigt war, wenn die Zweitaufgabe in Phase 2 stattfand, als wenn diese in Phase 1 stattfand. Diese Ergebnisse lassen erneut vermuten, dass vorrangig Aufmerksamkeitsprozesse, die in Phase 2 des Lern-Trials stattfinden, eine entscheidende Rolle für die Entstehung des Vortest-Effekts spielen.

9.3. Diskussion Experimente 8-9

Bei Betrachtung der vorliegenden Befunde zeigte sich für die Bedingung unter voller Aufmerksamkeit der typische Vortest-Effekt, mit einer besseren Erinnerungsleistung im finalen Cued-Recall-Test für Target-Items, die in der Vortest-Bedingung präsentiert wurden, verglichen mit Target-Items der Lern-Bedingung. Mit diesen Ergebnissen ließen sich die Befunde zahlreicher Studien replizieren, in denen unter Verwendung von schwach assoziierten Wortpaaren ein Vortest-Effekt gefunden werden konnte (z. B. Hays et al., 2013; Kliegl et al., 2024; Kornell, 2014). Unter geteilter Aufmerksamkeit ergab sich ein noch größerer Vortest-Effekt, wenn die Aufmerksamkeit während Phase 1 durch die Bearbeitung

einer Zweitaufgabe geteilt wurde, wohingegen der Vortest-Effekt ausblieb, wenn die Zweitaufgabe während Phase 2 eines Lern-Trials bearbeitet werden sollte. Die Ergebnisse zusätzlicher Analysen, die nur die Erinnerungsleistung der Vortest-Bedingung unter voller und unter geteilter Aufmerksamkeit berücksichtigten, erweiterten diese Befunde noch, indem sie zeigten, dass, verglichen mit der Vortest-Bedingung unter voller Aufmerksamkeit, in der Vortest-Bedingung mit geteilter Aufmerksamkeit eine Ablenkung sowohl in Phase 1 als auch in Phase 2 zu einer schlechteren Erinnerungsleistung an die Target-Wörter führte. Diese Beeinträchtigung war jedoch ausgeprägter, wenn die Ablenkung in Phase 2 stattfand. Dieses Ergebnismuster ergab sich unabhängig davon, ob die verwendete Zweitaufgabe materialallgemeine Interferenz (Experiment 8) oder materialspezifische Interferenz (Experiment 9) induzierte.

Die durchgeführten Experimente stellen erst die zweite Untersuchung des Einflusses von geteilter Aufmerksamkeit auf den Vortest-Effekt dar und sind die ersten Experimente, in denen gezielt und separat die Rolle von Aufmerksamkeit in Phase 1 und Phase 2 eines Lern-Trials auf die Größe des Vortest-Effekts betrachtet wurde. Dabei legen die Ergebnisse die Schlussfolgerung nahe, dass sowohl in Phase 1 des Lern-Trials, in der die Versuchspersonen das Lernmaterial entweder nur lernen (Lern-Bedingung) oder einen Vortest bearbeiten (Vortest-Bedingung) als auch in Phase 2 des Lern-Trials, in der in der Lern-Bedingung das Material weiter gelernt werden kann und in der Vortest-Bedingung das korrektive Feedback präsentiert wird, kognitive Prozesse stattfinden, die Aufmerksamkeitsressourcen erfordern. Dabei scheint aber vor allem Phase 2 besonders viele Aufmerksamkeitsressourcen zu beanspruchen. Daraus lässt sich ableiten, dass bei der Anwendung des Vortest-Paradigmas insbesondere in der Phase des korrektiven Feedbacks auf die Verfügbarkeit der vollen Aufmerksamkeitsressourcen und die Vermeidung von Ablenkung geachtet werden sollte, um den positiven Effekt von Vortests zu erhalten. Eine Ablenkung während des initialen Vortests selbst scheint hingegen nur in geringem Ausmaß durch eine Zweitaufgabe behindert zu werden und kann im Vergleich zu einer Lern-Bedingung (ebenfalls unter geteilter Aufmerksamkeit) sogar zu einem noch größeren Erinnerungsvorteil führen.

Die Ergebnisse, dass die Bearbeitung einer Zweitaufgabe während Phase 2 des Lern-Trials, nicht aber während Phase 1 des Lern-Trials zu einem Ausbleiben des Vortest-Effekts

führt, lassen vermuten, dass sich der Ursprung des Vortest-Effekts vorrangig in Phase 2 finden lässt, bzw. insbesondere in dieser Phase kontrollierte und aufmerksamkeitsressourcenfordernde Prozesse ablaufen, die für die Entstehung des Vortest-Effekts entscheidend zu sein scheinen.

Diese Schlussfolgerung lässt sich gut in Einklang mit der Aufmerksamkeits-Hypothese des Vortest-Effekts bringen. Denn diese geht davon aus, dass das Bearbeiten eines Vortests die Neugierde der Probanden steigert, die korrekte Lösung lernen zu wollen und somit das aufmerksame Enkodieren des in Phase 2 präsentierten Feedbacks fördert, was im finalen Test zu einer besseren Erinnerungsleistung für das vorgetestete Material führt. Diese Annahme passt gut zu einer Studie von Pan et al. (2020), in der die Versuchspersonen subjektiv einschätzen sollten, wie viel Aufmerksamkeit sie auf die Enkodierung mehrerer Lernvideos mit oder ohne die Bearbeitung eines Vortests verwendeten und wie sehr es dabei zu Ablenkungen kam. Hier wurde für alle Videosequenzen ein höheres Aufmerksamkeitsrating in der Vortest-Bedingung abgegeben als in der Lern-Bedingung. Die Präsentation einer Zweitaufgabe während Phase 2 sollte dementsprechend dazu führen, dass den Versuchspersonen weniger Aufmerksamkeitsressourcen zur Verfügung stehen und dadurch das aufmerksamere Enkodieren der Target-Items nach einem Vortest eingeschränkt ist, wodurch sich der Vorteil der Vortest-Bedingung gegenüber der Lern-Bedingung, also der Vortest-Effekt, reduzieren sollte. Genau dies konnte in den Experimenten 8b und 9b gezeigt werden.

Die zusätzlichen Analysen offenbarten, dass es, auch wenn die Zweitaufgabe während Phase 1 des Lern-Trials stattfand, zu einer geringen, aber signifikanten Verschlechterung der Erinnerungsleistung für Target-Wörter in der Vortest-Bedingung kam, verglichen mit den Target-Wörtern der Vortest-Bedingung unter voller Aufmerksamkeit (Experimente 8a und 9a). Diese Ergebnisse scheinen sich auf den ersten Blick nur schwer mit der Aufmerksamkeits-Hypothese erklären zu lassen. So lassen diese Ergebnisse darauf schließen, dass auch in Phase 1 aufmerksamkeitsressourcenfordernde Prozesse stattfinden, die durch eine Zweitaufgabe gestört werden und zu einer reduzierten Erinnerungsleistung in der Vortest-Bedingung führen. Eine mögliche Erklärung für diesen Befund stammt aus einer Vorgängerstudie, in der gezeigt werden konnte, dass unter voller Aufmerksamkeit die Aufmerksamkeit der Teilnehmer erst anstieg, *nachdem* sie einen Rateversuch kreierten (Potts

et al., 2019). Da die Versuchspersonen im Durchschnitt allerdings erst nach einigen Sekunden einen Rateversuch abgeben, sollte dieser Anstieg der Aufmerksamkeit erst in den letzten Sekunden von Phase 1 auftreten und sich dann in Phase 2 positiv auf die Effektivität der Enkodierung des korrektiven Feedbacks auswirken. Wenn nun aber die Aufmerksamkeit während Phase 1 durch eine Zweitaufgabe geteilt wird, sollte es nur in begrenztem Maße zu einem Anstieg der Aufmerksamkeit durch den Vortest kommen können, insbesondere dann, wenn der Rateversuch relativ früh in Phase 1 abgegeben wird, da die für den Anstieg der Aufmerksamkeit erforderlichen Ressourcen durch die Bearbeitung der Zweitaufgabe gebunden sein sollten. Dadurch könnte die effektivere Verarbeitung der korrekten Lösung in Phase 2 zumindest teilweise verhindert werden, weshalb es in der Vortest-Bedingung zu einer Beeinträchtigung der Erinnerungsleistung im finalen Test durch geteilte Aufmerksamkeit gekommen sein könnte, verglichen mit der Vortest-Bedingung unter voller Aufmerksamkeit, bei der die effektive Verarbeitung des korrekten Target-Items nicht gestört sein sollte.

Der Befund eines ausbleibenden Vortest-Effekts bei geteilter Aufmerksamkeit während Phase 2 ist ebenfalls konsistent mit den Annahmen der Error-Correction-Hypothese. Dieser Ansatz geht davon aus, dass es in der Vortest-Bedingung in Phase 2 zur Aktivierung von Fehlerkorrekturmechanismen kommt, durch die die Cue-Target-Verknüpfung gestärkt wird. Diese Stärkung führt in der Vortest-Bedingung im Vergleich zur Lern-Bedingung, in der dieser Prozess nicht stattfindet, zu einem späteren Erinnerungsvorteil. Da die Festigung der Verknüpfung zwischen dem Cue- und dem Target-Item kognitive Kontrolle und Aufmerksamkeit erfordert, sollte eine Reduktion der verfügbaren Aufmerksamkeitsressourcen während Phase 2 zu einer Reduktion des Vortest-Effektes führen. Dies entspricht den Ergebnissen der Experimente 8b und 9b. Die gefundene Beeinträchtigung der Erinnerungsleistung in der Vortest-Bedingung bei geteilter Aufmerksamkeit in Phase 1 des Lern-Trials (Experimente 8a und 9a) lässt sich hingegen nur schwer mit der Error-Correction-Hypothese vereinbaren.

Allerdings stellt sich die Frage, ob sich die vorliegenden Befunde möglicherweise auch noch mit anderen prominenten Erklärungsansätzen des Vortest-Effekts, wie der Elaborations-Hypothese oder der Search-Set-Hypothese, erklären lassen. Die Elaborations-Hypothese geht davon aus, dass es durch die Bearbeitung eines Vortests während Phase 1 zu

einer Aktivierung von Gedächtnisinhalten, die mit dem Cue-Wort verwandt sind, kommt und diese Gedächtnisinhalt anschließend mit dem Cue-Target-Paar verknüpft werden, wenn das korrektive Feedback in Phase 2 präsentiert wird. Diese zusätzliche Information kann dann wiederum in einem finalen Test als semantischer Mediator fungieren und den Abruf des Target-Items erleichtern (Huelser & Metcalfe, 2012; Cyr & Anderson, 2018). Allerdings ist unklar, inwiefern diese Elaborationsprozesse sowohl in Phase 1 als auch in Phase 2 Aufmerksamkeitsressourcen erfordern, da in der Literatur diskutiert wird, ob es sich bei Elaborationsprozessen um automatisierte Prozesse handelt, die dementsprechend nur wenig Aufmerksamkeitsressourcen erfordern (z. B. Anderson, 1983), oder ob es sich um anstrengende und für Ablenkung anfällige Prozesse handelt, die viele Aufmerksamkeitsressourcen erfordern (z. B. Craik et al., 1996, Mulligan, 2008). Dadurch lässt sich nur bedingt eine Aussage darüber treffen, inwiefern die vorliegenden Befunde zur Elaborations-Hypothese passen.

Die Search-Set-Hypothese hingegen geht davon aus, dass durch den Versuch, das korrekte Target-Wort in einem Vortest in Phase 1 zu erraten, ein sogenanntes Search-Set aktiviert wird. Dieses Search-Set beinhaltet mit dem Cue-Wort verwandte Inhalte, vor allem aber auch das korrekte Target-Wort, weshalb das korrekte Target-Wort effizienter enkodiert wird, wenn in Phase 2 das korrektive Feedback bereitgestellt wird, da das entsprechende Target-Wort bereits voraktiviert wurde (z. B. Grimaldi & Karpicke, 2012). Ähnlich wie bei der Elaborations-Hypothese ist auch bei der Search-Set-Hypothese allerdings unklar, ob es sich sowohl bei der Aktivierung des Search-Sets in Phase 1 als auch bei der Enkodierung des korrektiven Feedbacks in Phase 2 um Prozesse handelt, die Aufmerksamkeitsressourcen erfordern und somit anfällig für Ablenkung durch eine Zweitaufgabe sein sollten oder nicht. Daher ist auch bezüglich der Search-Set-Hypothese keine konkrete Aussage möglich, ob die vorliegenden Ergebnisse die Annahmen dieses Erklärungsansatzes stützen oder diesem widersprechen. Denkbar wäre aber beispielsweise, dass die Aktivierung des semantischen Netzwerks bzw. des Search-Sets während Phase 1 ausgehend vom Cue-Item einen automatisierten Prozess darstellt, der keine zusätzlichen Aufmerksamkeitsressourcen erfordert, wie das in der Literatur zur *spreading activation* (z. B. Balota & Lorch, 1986; Neely, 1977) angenommen wird. Hier sollte also eine Reduzierung der für die Lern-Aufgabe verfügbaren Aufmerksamkeitsressourcen in der Vortest-Bedingung nur zu einer geringeren

Beeinträchtigung führen. Die Integration der aktivierten Information in die Cue-Target-Verknüpfung bzw. die tiefere Enkodierung des im Search-Set voraktivierten Target-Wortes während Phase 2 hingegen könnte einen aktiven und kontrollierten Prozess darstellen, der Aufmerksamkeitsressourcen erfordert, weshalb eine Zweitaufgabe sich in dieser Phase negativ auf die Erinnerungsleistung ausgewirkt haben könnte.

Die Kontext-Hypothese erklärt den Vortest-Effekt durch eine größere Anzahl hilfreicher Kontext-Merkmale, die im finalen Test für Items der Vortest-Bedingung verglichen mit Items der Lern-Bedingung zur Verfügung stehen und dadurch den Abruf der Zielinformation erleichtern. Da die Integration der räumlich-zeitlichen Kontextdetails in das Cue-Target-Paar überwiegend automatisiert und unbewusst abläuft, sollte sie nicht sensibel für geteilte Aufmerksamkeit durch eine Zweitaufgabe sein. Vor diesem Hintergrund lässt sich dieser Erklärungsansatz anhand der beiden Experimente 8 und 9 nicht konkret überprüfen.

Insgesamt lassen sich die vorliegenden Ergebnisse gut in die Beobachtungen aus früheren Studien, die die Auswirkungen geteilter Aufmerksamkeit auf den klassischen Test-Effekt untersuchten, einordnen. So führten Buchin und Mulligan (2017, Experimente 2 und 3) beispielsweise eine typische Test-Effekt-Aufgabe durch, wobei die Versuchspersonen in Phase 1 des Lern-Trials zunächst Wortpaare (z. B. Stern – Nacht) lernen sollten. In Phase 2 des Lern-Trials konnte das Lernmaterial entweder wiederholt gelernt werden, oder es wurde ein interpolierter Test absolviert, bei dem das Cue-Wort präsentiert wurde und das dazugehörige Target-Wort abgerufen werden sollte (z. B. Stern – ?). Unter voller Aufmerksamkeit zeigte sich der klassische Test-Effekt, d. h. Wortpaare, die getestet wurden, wurden in einem späteren finalen Test besser erinnert als Wortpaare, die erneut gelernt werden sollten. Der Test-Effekt war jedoch sogar noch ausgeprägter, wenn in Phase 2 des Lern-Trials eine Zweitaufgabe durchgeführt wurde (für ähnliche Ergebnisse siehe Mulligan & Picklesimer, 2016; Buchin & Mulligan, 2019). Dieses Muster trat, wie in den vorliegenden Experimenten, unabhängig davon auf, ob materialallgemeine oder materialspezifische Interferenz induziert wurde.

Die Befunde der Experimente 8 und 9 lassen sich allerdings nicht nur in die Literatur zum klassischen Test-Effekt einordnen, sondern auch mit den Ergebnissen der einzigen, kürzlich durchgeführten Studie zur Untersuchung des Vortest-Effekts unter geteilter

Aufmerksamkeit vergleichen. In ihrer Studie teilten Mulligan und Buchin (2024), im Gegensatz zur vorliegenden Studie, die Aufmerksamkeit während der gesamten Vortest- und Lern-Trials auf, also sowohl in Phase 1 als auch in Phase 2 eines Trials, wobei die Ergebnisse zeigten, dass der Vortest-Effekt trotz dieser Ablenkung durch eine Zweitaufgabe bestehen blieb. Numerisch betrachtet war die Größe des Vortest-Effekts bezüglich der Hit-Rate unter geteilter Aufmerksamkeit sogar größer als unter voller Aufmerksamkeit (17 % vs. 11 %), was eine Parallele zu den Ergebnissen der aktuellen Experimente 8a und 9a darstellt.

Die Ergebnisse der vorliegenden Experimente 8b und 9b scheinen allerdings nur schwer mit den Befunden von Mulligan und Buchin (2024) vereinbar zu sein, denn die Tatsache, dass in dieser Studie der Vortest-Effekt bestehen blieb, obwohl während des gesamten Lern-Trials und damit auch während Phase 2 die Aufmerksamkeit geteilt wurde, könnte darauf hindeuten, dass Aufmerksamkeitsprozesse auch in Phase 2 des Durchgangs keine entscheidende Rolle spielen. Ein möglicher Faktor für die unterschiedlichen Ergebnisse könnte darin liegen, dass in der Studie von Mulligan und Buchin (2024) der finale Test als Rekognitionstest durchgeführt wurde, während in der vorliegenden Studie ein Cued-Recall-Format für den finalen Test gewählt wurde. Denn frühere Studien deuten darauf hin, dass der Vortest-Effekt in Rekognitionstests manchmal leichter auftritt als in Cued-Recall-Tests (z. B. Potts & Shanks, 2014; Potts et al., 2019; Seabrooke et al., 2021).

Auf einer allgemeineren Ebene stimmen die vorliegenden Ergebnisse jedoch mit zahlreichen früheren Befunden überein, die zeigen, dass geteilte Aufmerksamkeit die spätere Erinnerungsleistung stärker beeinträchtigt, wenn diese während des Enkodierens statt während des Abrufs der Lerninhalte stattfindet (für eine Übersicht siehe Mulligan, 2008). So ergab sich in den vorliegenden Experimenten 8 und 9 eine schlechtere Erinnerungsleistung unter geteilter als unter voller Aufmerksamkeit. Dies galt einerseits für die Lern-Bedingung verglichen mit der Vortest-Bedingung, wenn die Aufmerksamkeit in Phase 1 geteilt war (siehe Abbildung 7 und 10), und andererseits für die Vortest-Bedingung im Vergleich zur Lern-Bedingung, wenn die Aufmerksamkeit während Phase 2 geteilt war (Abbildung 8 und 11). Das erste Ergebnis reiht sich gut in frühere Studien ein, die zeigen, dass geteilte Aufmerksamkeit die Enkodierungsprozesse stärker beeinträchtigen kann als Abrufprozesse. So findet während Phase 1 in der Lern-Bedingung die Präsentation und die Enkodierung des Lernmaterials statt, welche sensibel für Ablenkung durch eine Zweitaufgabe sein sollte. In

der Vortest-Bedingung hingegen soll in Phase 1 der Vortest bearbeitet werden, was einen Abrufprozess darstellt, der weniger durch geteilte Aufmerksamkeit gestört werden sollte. Dadurch ergibt sich folglich eine Vergrößerung des Vortest-Effekts unter geteilter Aufmerksamkeit in dieser Phase. In Phase 2 hingegen findet in der Vortest-Bedingung die Enkodierung des Target-Wortes statt, das erst in dieser Phase in Form des korrektiven Feedbacks präsentiert und gelernt wird. In der Lern-Bedingung konnte das Wortpaar zu diesem Zeitpunkt bereits für 6 s gelernt und enkodiert werden, weshalb der Enkodierprozess hier möglicherweise bereits abgeschlossen oder nicht mehr anfällig für Ablenkung durch eine Zweitaufgabe ist. Folglich sollte die Manipulation geteilter Aufmerksamkeit in Phase 2 eine kleinere Auswirkung auf die Erinnerungsleistung in der Lern-Bedingung haben als auf die Erinnerungsleistung in der Vortest-Bedingung. Dadurch sollte sich ein reduzierter oder ausbleibender Vortest-Effekt zeigen, wenn die Aufmerksamkeit in dieser Phase geteilt wird. Genau das zeigen die Ergebnisse der vorliegenden Experimente.

Zusammenfassend lässt sich also festhalten, dass die Experimente 8 und 9 erste Hinweise über den Einfluss geteilter Aufmerksamkeit auf die Größe des Vortest-Effekts liefern konnten. Vor allem aber ergaben sich dadurch weitere wichtige Erkenntnisse für ein besseres Verständnis der zugrundeliegenden kognitiven Mechanismen, insbesondere über die Rolle von aufmerksamkeitsressourcenfordernden Prozessen für die Entstehung des Vortest-Effekts. So konnte gezeigt werden, dass der Vortest-Effekt nur dann ausbleibt, wenn eine Zweitaufgabe während Phase 2, in der das korrektive Feedback präsentiert wird, stattfindet, nicht aber, wenn die Zweitaufgabe in Phase 1 bearbeitet werden soll. Insgesamt deutet dies folglich darauf hin, dass die Prozesse, die in Phase 2 des Lern-Trials ablaufen, mehr Aufmerksamkeitsressourcen erfordern als die Prozesse, die in Phase 1 ablaufen. Diese Befunde stehen vor allem im Einklang mit der Aufmerksamkeits-Hypothese des Vortest-Effekts, widersprechen aber nicht direkt anderen prominenten Erklärungsansätzen.

10. Gesamtdiskussion

10.1. Zusammenfassung der Befunde

Obwohl vor allem beginnend mit den Studien von Kornell et al. (2009) eine zweite Forschungswelle zur Untersuchung des Vortest-Effekts in Gang kam und seither eine Vielzahl von Studien sich mit diesem Gedächtniseffekt auseinandersetzten, bleiben bis zum heutigen Tage noch viele Forschungsfragen offen und viele relevante Aspekte über diesen Effekt unerforscht. So finden sich in der bisherigen Literatur beispielsweise noch keine Untersuchung darüber, inwiefern sich die Anzahl an Vortests auf die Größe des Effekts auswirkt, wohingegen dieser empirische Aspekt für andere Arten der Abrufübung bereits vielfach betrachtet wurde. Daher wurde in den Experimenten 1-4 der Einfluss wiederholter Rateversuche während des Vortests auf die spätere Erinnerungsleistung untersucht. Dabei zeigte sich, dass sich die Durchführung wiederholter Vortests noch positiver auf den Abruf der Zielinformationen auswirkt als nur ein einziger Vortest, und dass die Anzahl an Vortests somit einen wichtigen Einflussfaktor auf die Größe dieses Effekts darstellt. Sowohl für zwei Vortests als auch für drei Vortests konnte mit jedem zusätzlichen Testdurchgang eine Zunahme des Vortest-Effekts beobachtet werden. Dieser Befund ergab sich unabhängig davon, ob die Gesamtlänge eines Trials mit steigender Anzahl an Rateversuchen zunahm (Experimente 1, 2, 4) oder konstant gehalten wurde (Experiment 3). Zudem ergab sich dieses Ergebnismuster sowohl für einfache Cue-Target-Paare (Experimente 1 und 3) als auch für komplexeres Lernmaterial wie Swahili-Deutsch-Vokabeln (Experiment 2) und Textpassagen (Experiment 4).

Zudem war bis zum Zeitpunkt der vorliegenden Arbeit noch ungeklärt, ob sich der Vortest-Effekt auch durch inzidentelles Lernen induzieren lässt und ob sich dieser Effekt auch dann zeigt, wenn im finalen Test neue, semantisch verwandte Cue-Items anstelle der originalen Cue-Items als Abrufreize verwendet werden. Die Experimente 5 bis 7 offenbarten, dass sich auch dann ein Erinnerungsvorteil für Target-Items der Vortest-Bedingung einstellt, wenn der Lernprozess unbewusst und ohne unmittelbare Lerninstruktion stattfand. Außerdem zeigten die gefundenen Ergebnisse, dass die Bearbeitung eines Vortests einerseits

zu einer besseren Erinnerungsleistung für die gelernten Cue- und Target-Items, andererseits aber auch zu einem höheren Anteil an False Alarms für verwandtes Material, wie semantische Mediatoren (Experiment 5), führt. Wenn die semantischen Mediatoren jedoch als Hinweisreiz in einem finalen Cued-Recall-Test anstelle des ursprünglich gelernten Cue-Items präsentiert wurden (Experiment 7), konnte ein Vortest-Effekt mit einer besseren Erinnerungsleistung für gelernte Target-Items in der Vortest-Bedingung gefunden werden. Diese Ergebnisse offenbarten zum ersten Mal, dass auch die Verwendung semantisch verwandter Information als Cue-Items zu einem Vortest-Effekt führen kann.

Die abschließend durchgeführten Experimente 8 und 9 widmeten sich der Frage, wie geteilte Aufmerksamkeit in den beiden Phasen des Lern-Trials den Vortest-Effekt beeinflusst. Die Befunde dieser Experimente sollten Aufschluss darüber geben, in welcher der beiden Phasen vorrangig Aufmerksamkeitsressourcen erforderlich sind, um einen Erinnerungsvorteil in der Vortest-Bedingung zu ermöglichen. Dabei liefern die Ergebnisse erste Hinweise darauf, dass sich die Verfügbarkeit von Aufmerksamkeitsressourcen in diesen beiden Phasen unterschiedlich auf den Vortest-Effekt auswirkt. Während geteilte Aufmerksamkeit in Phase 2 zu einem Ausbleiben des Effekts führte (Experimente 8b und 9b), kam es bei geteilter Aufmerksamkeit in Phase 1 hingegen zu einem noch prominenteren Vortest-Effekt (Experimente 8a und 9a). Mit den Ergebnissen der neun durchgeführten Experimente konnte insgesamt also ein zentrales Ziel der Arbeit erreicht werden, das in der Erweiterung der Vortest-Literatur und der gezielten Betrachtung noch offener Fragestellungen zur Schließung bestehender empirischer Lücken bezüglich des Vortest-Paradigmas bestand.

10.2. Theoretische Schlussfolgerungen

Das zweite Hauptziel der Arbeit war es, grundlegende Aussagen über mögliche kognitive Mechanismen des Vortest-Effekts treffen zu können und konkrete Annahmen einzelner Erklärungsansätze zu überprüfen, um somit die theoretischen Hintergründe dieses Effekts besser verstehen zu können.

Einen prominenten Erklärungsansatz des Vortest-Effekts stellt die Elaborations-Hypothese des Vortest-Effekts dar (Carpenter, 2009, 2011). Dieser Erklärungsansatz geht davon aus, dass durch die Bearbeitung eines Vortests Informationen aktiviert werden, die

semantisch mit dem Cue verwandt sind, und dass diese Informationen das Cue-Item mit dem dazugehörigen Target-Item verknüpfen und somit als Mediatoren fungieren (James, 2022). Dadurch entstehen mehrere mögliche Abrufwegen, um bei der Präsentation des Cue-Items Zugang zur korrekten Zielinformation zu erhalten. Diese Annahmen basieren auf dem Prinzip der *spreading activation* (Anderson, 1983; Collins & Loftus, 1975), demzufolge es ausgehend von einem sogenannten Knotenpunkt, in diesem Fall dem Cue-Item, zu einer Ausbreitung der Gedächtnisaktivierung auf nahe verknüpfte Knotenpunkte, in diesem Fall den potenziellen Mediatoren, kommt. Diese Mediatoren wiederum erleichtern in einem finalen Test den Abruf der Target-Information, wodurch sich eine bessere Erinnerungsleistung in der Vortest-Bedingung zeigt als in der Lern-Bedingung, in der diese zusätzlichen Knotenpunkte, die gleichbedeutend mit zusätzlichen Abrufwegen verstanden werden können, fehlen.

Die Ergebnisse der Experimente 1-4 bestätigen diese Annahmen nicht nur, sondern deuten auch darauf hin, dass die Aktivierung dieser naheliegenden Knotenpunkte mit jedem weiteren Vortest und generierten Rateversuch weiter zunimmt, und dass das semantische Netzwerk damit noch detailreicher wird. Mit jedem zusätzlichen Rateversuch entstehen somit mehr potenzielle Abrufwegen vom Cue-Item zum dazugehörigen Target-Item, die in einem späteren finalen Test zur Verfügung stehen. Dadurch erhöht sich die Wahrscheinlichkeit, dass eine dieser Routen im finalen Test genutzt wird, um die korrekte Zielinformation abzurufen, wodurch der Erinnerungsvorteil für diese Target-Items nach wiederholten Vortests noch ausgeprägter ist als nach nur einem Rateversuch.

Außerdem deuten die Befunde der durchgeführten Experimente 5 bis 7 darauf hin, dass die Aktivierung der semantisch verwandten Information auch dazu führt, dass diese Informationen als vertrauter eingeschätzt werden. So wurden diese semantischen Mediatoren in der Vortest-Bedingung häufiger fehlerhaft als bekannt bzw. gelernt eingeschätzt als in der Lern-Bedingung. Dies könnte darauf hindeuten, dass in der Lern-Bedingung durch die ausbleibende Voraktivierung eine bessere Unterscheidung zwischen gelernten Cue- und Target-Items und verwandten Mediatoren möglich ist als in der Vortest-Bedingung. Denn in der Vortest-Bedingung könnten sowohl die Cue- und Target-Items als auch die semantischen Mediatoren (aufgrund ihrer Verknüpfung mit dem Cue-Target-Paar) als bekannt abgespeichert worden sein. Dies könnte im finalen Rekognitionstest dazu geführt haben, dass

gelernte alte und ungelernte neue Items schlechter voneinander unterschieden werden konnten. Durch die Elaborationsprozesse könnte es folglich zu dem unerwünschten negativen Transfereffekt einer größeren Anzahl an Fehleinschätzungen bezüglich semantischer Mediatoren gekommen sein.

Allerdings führen die Elaborationsprozesse nicht nur zu negativen, sondern auch zu positiven Transfereffekten bezüglich der semantischen Mediatoren. So konnten die Ergebnisse aus Experiment 7 zeigen, dass die Bearbeitung eines Vortests, wie von der Elaborations-Hypothese angenommen, zu einer Aktivierung verwandter Information und der Integration dieser Informationen in das Cue-Target-Paar führt (z. B. Grimaldi & Karpicke, 2012; Huelser & Metcalfe, 2012). Diese semantischen Mediatoren werden dann als zusätzliche Hinweisreize im finalen Test reaktiviert und erleichtern so als „Sprungbrett“ den Zugang vom Cue-Item zum korrekten Target-Item. Außerdem scheinen semantische Mediatoren nicht nur ein Hilfsmittel darzustellen, um die Abrufroute vom Cue-Item zum Target-Item zu vereinfachen, sondern sie scheinen durch ihre Verknüpfung mit dem Target-Item auch selbst als effektive Cues zu fungieren und so zu einem korrekten Abruf der Target-Information zu führen. Denn sowohl, wenn der originale Cue, als auch, wenn ein semantischer Mediator als Cue-Item verwendet wurde, konnte in der Vortest-Bedingung ein Erinnerungsvorteil für die Target-Information gefunden werden.

Insgesamt lassen sich durch die Ergebnisse der Experimente 1-7 (mit Ausnahme von Experiment 2) zentrale Annahmen der Elaborations-Hypothese bestätigen, wodurch die Bedeutung von semantischen Elaborationsprozessen für die Entstehung des Vortest-Effekts unterstrichen und teilweise noch präzisiert werden konnte. Die in den Experimenten 8 und 9 gezeigten Befunde stehen zwar nicht im Widerspruch zu den Annahmen der Elaborations-Hypothese, allerdings lassen sich aus diesen Experimenten keine konkreten Aussagen über die Rolle dieses Erklärungsansatzes für den Vortest-Effekt ableiten (siehe Tabelle 8).

Einen zweiten wichtigen Mechanismus, der auf Grundlage der vorliegenden Ergebnisse eine wichtige Rolle für die Entstehung des Vortest-Effekts zu spielen scheint, stellen Aufmerksamkeitsprozesse dar. So sind die gezeigten Befunde der Experimente 1 bis 4 und vor allem der Experimente 8 und 9 konsistent mit den Annahmen der Aufmerksamkeits-Hypothese des Vortest-Effekts (Potts & Shanks, 2014). Dieser Erklärungsansatz geht davon aus, dass die Bearbeitung eines Vortests den Lernenden eine

Diskrepanz zwischen dem aktuellen Wissenstand und der abgefragten Information aufzeigt, was zu einer gesteigerten Neugier und Motivation führt, diese Lücke zu schließen und die korrekte Lösung zu erfahren (z. B. Zawadzka & Hanczakowski, 2019). Dafür wird der Aufmerksamkeitsfokus vor allem auf die Phase 2 zentriert, in der das korrektive Feedback präsentiert und aufgrund der zusätzlichen Aufmerksamkeitsressourcen effektiver enkodiert wird. Diese verbesserte Enkodierung der Target-Information wiederum führt in einem späteren finalen Test zu einer besseren Erinnerungsleistung für diese Items in der Vortest-Bedingung verglichen mit der Lern-Bedingung, in der dieser gezielte Aufmerksamkeitsfokus nicht stattfindet (z. B. Seabrooke et al., 2021). Auch in den vorliegenden Experimenten könnte die Bearbeitung eines Vortests zu der Feststellung einer Wissenslücke und der anschließend besseren Enkodierung des korrektiven Feedbacks, um diese Wissenslücke zu schließen, geführt haben, wobei insbesondere die Ergebnisse der Experimente 8 und 9 sowie die der Experimente 1-4 diese Annahme zu bestätigen und die Aufmerksamkeits-Hypothese noch zu erweitern scheinen.

So deuten die Befunde darauf hin, dass das Erkennen einer Diskrepanz zwischen dem eigenen Kenntnisstand und der gesuchten Zielinformation während des Vortests in Phase 1 des Lern-Trials nur wenige Aufmerksamkeitsressourcen zu erfordern und somit einen (zumindest teilweise) automatisierten Prozess darzustellen scheint. So wirkte sich die Störung durch eine Zweitaufgabe in dieser Phase kaum negativ auf die spätere Erinnerungsleistung aus, was dafür sprechen könnte, dass es trotz der Ablenkung in Phase 1 zu einer Konzentration der Aufmerksamkeitsressourcen auf die Enkodierung der Target-Information in Phase 2 gekommen sein könnte, die zu einem Vortest-Effekt führte. Die effektive Enkodierung und aufmerksame Verarbeitung der Target-Items in Phase 2 hingegen scheint einen aktiven und kontrollierten Prozess darzustellen, der Aufmerksamkeitsressourcen erfordert. Denn eine Reduktion der in dieser Phase 2 verfügbaren Aufmerksamkeitsressourcen durch eine Zweitaufgabe führte zu einem Ausbleiben des Vortest-Effekts. Kann also die korrekte Target-Information in Phase 2 durch eine Störung nicht aufmerksamer und effektiver enkodiert werden, kommt es zu vergleichbaren Erinnerungsleistungen in der Vortest- und der Lern-Bedingung, was die Bedeutung dieses Prozesses für die Entstehung des Vortest-Effekts betont. Darüber hinaus scheinen die Neugier und Motivation, die korrekte Lösung zu lernen, mit zunehmender

Anzahl an Vortests und generierten Rateversuchen noch weiter zu steigen. Dies könnte zu einer noch stärkeren Fokussierung der Aufmerksamkeitsressourcen und einer noch effektiveren Verarbeitung der Target-Items führen, wenn diese in Form von korrektivem Feedback präsentiert werden. Damit ließe sich erklären, warum es im finalen Test zu einem noch besseren Abruf der Target-Information und einem noch größeren Vortest-Effekt nach wiederholten Vortests verglichen mit nur einem Vortest gekommen ist (Experimente 1-4).

Zusammenfassend deuten insbesondere die Ergebnisse der Experimente 1 bis 4 und die Experimente 8b und 9b darauf hin, dass an der Entstehung des Vortest-Effekts Prozesse beteiligt sind, die Aufmerksamkeitsressourcen erfordern, wodurch sich die Annahmen der Aufmerksamkeits-Hypothese bestätigen ließen. Weniger konkrete Rückschlüsse bezüglich dieses Erklärungsansatzes lassen sich hingegen aus den Experimenten 5-7 ziehen, da die Aufmerksamkeits-Hypothese keine direkten Annahmen bezüglich der Rolle semantischer Mediatoren für den Vortest-Effekt trifft (siehe Tabelle 8).

Tabelle 8

Theoretische Einordnung der Befunde der durchgeführten Experimente

Erklärungsansätze	Die Ergebnisse der folgenden Experimente...		
	...lassen sich durch Hypothese erklären	...lassen sich nur schwer durch Hypothese erklären/widersprechen Hypothese	...erlauben keine konkrete Aussage über Hypothese
Elaborations-Hypothese	Experimente 1, 3-7	Experiment 2	Experimente 8, 9
Aufmerksamkeits-Hypothese	Experimente 1-4, 8b, 9b	Experimente 8a, 9a	Experimente 5-7
Error-Correction-Hypothese	Experimente 8b, 9b	Experimente 1-7, 8a, 9a	-
Kontext-Hypothese	Experimente 1-4	-	Experimente 5-9
Search-Set-Hypothese	Experimente 1-4	-	Experimente 5-9
Cue-Overload-Prinzip	-	Experimente 1-4	-

Anmerkung: Überblick, ob die Ergebnisse der unterschiedlichen Experimente mit den Annahmen der verschiedenen Erklärungsansätze übereinstimmen, diesen (teilweise) widersprechen oder keine direkte Schlussfolgerung ermöglichen.

Die beiden genannten Erklärungsansätze schließen sich allerdings nicht zwingend gegenseitig aus, sondern es wäre auch möglich, dass mehrere kognitive Mechanismen oder

Prozesse zu den beobachteten Effekten beitragen. Einen neuen Erklärungsansatz, der möglicherweise eine Verknüpfung von Aufmerksamkeits- und Elaborationsprozessen beinhalten könnte, ist die sogenannte Aufmerksamkeitsfenster-Hypothese von Sana und Carpenter (2023). Dieser Erklärungsansatz geht davon aus, dass es durch die Bearbeitung eines Vortests zur Öffnung eines Aufmerksamkeitsfensters kommt, in dem ein Suchprozess nach der korrekten Lösung für das Vortest-Item stattfindet. Während dieses Suchprozesses kommt es zu einer Steigerung der verwendeten Aufmerksamkeitsressourcen, bis die Präsentation der korrekten Lösung zur Schließung des Aufmerksamkeitsfensters und einer Reduktion der aufgewendeten Aufmerksamkeitsressourcen führt. So wäre es denkbar, dass das Feststellen einer Wissenslücke während des Vortests zu einer gesteigerten Neugierde für die korrekte Lösung beiträgt, wie dies die Aufmerksamkeits-Hypothese annimmt, und dass es dadurch zu der Öffnung des Aufmerksamkeitsfensters kommt. Die zusätzlichen Aufmerksamkeitsressourcen, die anschließend aktiviert werden, könnten wiederum während des Suchprozesses nach der korrekten Lösung für die Aktivierung semantisch verwandter Information und die Integration der korrekten Lösung in das voraktivierte semantische Netz, im Sinne der Elaborations-Hypothese, genutzt werden. Diese Hypothese könnte also eine Erklärung darstellen, die Aufmerksamkeits- und Elaborationsprozesse nicht als exklusive, eigenständige Mechanismen betrachtet, sondern einen Zusammenhang zwischen diesen beiden Prozessen vermuten lässt.

Einen anderen Erklärungsansatz für den Vortest-Effekt stellt die Error-Correction-Hypothese dar (Carrier & Pashler, 1992). Dieser Ansatz nimmt an, dass ein fehlerhafter Rateversuch unmittelbar gefolgt von korrektivem Feedback ein Fehlersignal auslöst, durch das Fehlerkorrekturmechanismen aktiviert werden. Diese Mechanismen erfordern kognitive Kontrolle und Aufmerksamkeit, weshalb es durch eine Beeinträchtigung dieser Prozesse, z. B. durch eine Zweitaufgabe, zu einem Ausbleiben des Vortest-Effekts kommen sollte. Die Ergebnisse der Experimente 8b und 9b, die genau dieses Ergebnismuster zeigen, lassen sich daher gut mit der Error-Correction-Hypothese erklären. Allerdings widersprechen vor allem die Befunde der Experimente 5 bis 7 den Grundannahmen der Error-Correction-Hypothese. Gemäß diesem Erklärungsansatz sollte es durch die Fehlerkorrekturmechanismen zu einer Stärkung der Cue-Target-Verknüpfung und einer Inhibition aller weiteren Verknüpfungen mit dem Cue-Item kommen. Folglich sollte ausschließlich das ursprüngliche Cue-Item zu

einem korrekten Abruf des Target-Items führen, nicht aber die Präsentation eines semantischen Mediators, wie dies in den Experimenten 5-7 der Fall war. Da aber vor allem diese Experimente die Error-Correction-Hypothese gezielt überprüften, sollte die Übereinstimmung dieses Erklärungsansatzes mit den Ergebnissen der Experimente 8b und 9b mit Vorsicht interpretiert werden. Insgesamt deutet die vorliegende Arbeit also darauf hin, dass Fehlerkorrekturmechanismen für die Entstehung des Vortest-Effekts eine eher untergeordnete Rolle spielen, was wiederum zu anderen Studien passt, die ebenfalls den Annahmen der Error-Correction-Hypothese widersprechen (z. B. Seabrooke et al., 2021).

Wie in Tabelle 8 dargestellt ist, lassen sich die Ergebnisse der Experimente 1-4 nur schwer mit den Annahmen des Cue-Overload-Prinzips erklären. Gemäß diesem Ansatz sollte ein Rateversuch mit dem Cue-Item verknüpft werden, wodurch dieser in einem späteren finalen Test eine Konkurrenz zum korrekten Target-Item darstellt und den Abruf dieses Target-Items erschwert. Wiederholte Rateversuche sollten demnach eine noch größere Konkurrenzsituation verursachen und zu einer schlechteren Erinnerungsleistung führen als nur ein Rateversuch. Über alle vier Experimente hinweg zeigte sich hingegen konsistent ein Anstieg der Erinnerungsleistung und eine Zunahme des Vortest-Effekts mit steigender Anzahl an Rateversuchen. Die Ergebnisse deuten also darauf hin, dass zumindest bei bis zu drei Rateversuchen während des Vortests keine hinderliche Überladung des Cue-Items auftritt, wodurch die Annahme des Cue-Overload-Prinzips nicht bestätigt werden konnte.

Über die Bedeutung von Kontext-Effekten und der Aktivierung eines Search-Sets für den Vortest-Effekt lassen sich auf Grundlage der vorliegenden Ergebnisse nur vereinzelt erste Rückschlüsse ziehen. So passen die Ergebnisse der Experimente 1-4 auch zu den Annahmen dieser beiden Hypothesen. Denn sowohl die Vielfalt an Kontextmerkmalen (Kontext-Hypothese) als auch die Wahrscheinlichkeit, dass das korrekte Target-Item während des Vortests in einem Search-Set bereits aktiviert und daher besser enkodiert wird (Search-Set-Hypothese), sollte mit steigender Anzahl an Rateversuchen während des Vortests zunehmen. Da es insbesondere diese beiden Prozesse sind, die gemäß den Erklärungsansätzen den Erinnerungsvorteil in der Vortest-Bedingung hervorrufen, lässt sich ein größerer Vortest-Effekt nach wiederholten Rateversuchen verglichen mit nur einem Rateversuch gut mit den Annahmen der beiden Hypothesen vereinbaren. Allerdings lassen sich auf Grundlage dieser beiden Theorien keine konkreten Erwartungen bezüglich der Rolle

semantischer Mediatoren (Experimente 5-7) und geteilter Aufmerksamkeit (Experimente 8 und 9) treffen, weshalb insgesamt der Einfluss dieser beiden Mechanismen für den Vortest-Effekt anhand der vorliegenden Arbeit nur bedingt beurteilt werden kann (siehe Tabelle 8).

Zuletzt wäre auch die Übertragung des Prinzips der *desirable difficulties*, das von Bjork (1994) eingeführt wurde, auf den Vortest-Effekt denkbar. Dieses Prinzip geht davon aus, dass es durch das Schaffen einer schwierigeren Lernumgebung später zu besseren Erinnerungsleistungen kommt. Dabei konnte bereits gezeigt werden, dass das Generieren einer eigenen Antwort, anstelle der direkten Darbietung der Lösung, eine derartige erwünschte Schwierigkeit darstellen könnte (Bjork & Bjork, 2011). Zudem konnten Potts und Shanks (2014) in ihrer Studie zeigen, dass Lernende stets niedrigere JOLs für vorgetestete Items abgaben als für gelernte Items, d. h., dass es als unwahrscheinlicher eingeschätzt wurde, diese Items in einem späteren Test korrekt abrufen zu können. Dies könnte widerspiegeln, dass diese Items von den Versuchspersonen als schwieriger zu lernen betrachtet wurden und dass diese Schwierigkeit zu einer besseren Langzeiterinnerung in Form des Vortest-Effekts beiträgt. Insgesamt lässt sich festhalten, dass durch die berichteten Ergebnisse wichtige Erkenntnisse über mögliche kognitive Mechanismen des Vortest-Effekts gewonnen werden konnten, es aber in Zukunft noch weiterer Forschung bedarf, um diese Prozesse noch besser verstehen und die Bedeutung weiterer Erklärungsansätze konkret überprüfen zu können.

10.3. Einordnung der Befunde in die bisherige Literatur

Insgesamt ließ sich, mit Ausnahme von Experiment 6, in allen Experimenten der typische Vortest-Effekt finden. Es zeigte sich dabei in einem späteren finalen Test jeweils eine bessere Erinnerungsleistung in der Vortest-Bedingung verglichen mit der Lern-Bedingung, unabhängig davon, ob im finalen Test ein Rekognitionstest (Experiment 5) oder ein Cued-Recall-Test (alle weiteren Experimente) verwendet wurde. Des Weiteren ergab sich dieser Erinnerungsvorteil in der Vortest-Bedingung auch für unterschiedlich komplexes Lernmaterial, von einfacheren Wortpaaren (z. B. Experimente 1 und 8) bis hin zu Vokabeln (Experiment 2) oder Prosatext-Passagen (Experiment 4).

Mit den durchgeführten Experimenten konnten also die Befunde vorheriger Studien, die den Vortest-Effekt für unterschiedliches Lernmaterial (z. B. Kornell et al., 2014; Potts &

Shanks, 2014; Richland et al., 2009) und unterschiedliche Testformate (z. B. Kliegl et al., 2024; Seabrooke et al., 2019) zeigen konnte, repliziert werden. Außerdem liefern die Ergebnisse der Experimente 8a und 9a den ersten Hinweis darauf, dass die Größe des Erinnerungsvorteil durch einen Vortest unter geteilter Aufmerksamkeit sogar noch zunimmt, ähnlich wie sich dies bereits in Studien zum klassischen Test-Effekt finden ließ (z. B. Buchin & Mulligan, 2019; Mulligan & Buchin 2022a).

Die Befunde der durchgeführten Experimente legen also bezüglich des Einflusses von geteilter Aufmerksamkeit auf die Größe des Gedächtniseffekts eine Parallele zwischen dem klassischen Test-Effekt und dem Vortest-Effekt offen. Diese Ergebnisse stellen eine Erweiterung der bereits gefundenen Gemeinsamkeiten zwischen dem klassischen Test-Effekt und dem Vortest-Effekt dar. So konnte bisher schon gezeigt werden, dass durch die Anwendung beider Paradigmen zeitabhängiges Vergessen reduziert werden kann. Zeitabhängiges Vergessen beschreibt dabei den typischen Befund, dass die Erinnerungsleistung für gelernte Informationen mit zunehmendem zeitlichem Abstand zwischen dem Lernen und dem Abruf der Information abnimmt, wie dies auch in der klassischen Vergessenskurve von Ebbinghaus (1885) dargestellt ist (für einen Überblick siehe Rubin & Wenzel, 1996). Die Vergrößerung des Retentionsintervalls zwischen der Lern-Phase und dem finalen Test führt also typischerweise zu einer Abnahme der korrekt erinnerten Items. Diese Abnahme der Erinnerungsleistung fällt aber durch die Bearbeitung eines Tests entweder nach dem Lernen (z. B. Carpenter et al., 2008; Larsen et al., 2013; für Metaanalyse siehe Adesope et al., 2017) oder bereits vor dem Lernen (z. B. Kliegl et al., 2024) deutlich geringer aus, als wenn das Material nur oder wiederholt gelernt wird. Dadurch ergibt sich mit wachsendem Retentionsintervall häufig auch eine Zunahme des klassischen Test-Effekts bzw. des Vortest-Effekts. Außerdem konnten in der bisherigen Literatur Hinweise darauf gefunden werden, dass sich auch Vergessen durch retroaktive Interferenz mittels einer Abrufübung im klassischen Test-Effekt-Paradigma (Halamish & Bjork, 2011; Potts & Shanks, 2012) wie auch im Vortest-Paradigma (Kliegl et al., 2023) reduzieren lässt.

Eine zusätzliche Ergänzung der bisher gefundenen Parallelen zwischen dem klassischen Test-Effekt und dem Vortest-Effekt stellt zudem der Befund dar, dass nicht nur die wiederholte Bearbeitung von Tests nach dem Lernen zu einem noch größeren Erinnerungsvorteil und einem Anstieg des Test-Effekts führen kann (siehe Yang et al., 2021

für Metaanalyse), sondern wiederholt durchgeführte Tests, wie in den Experimenten 1-4 gezeigt, auch zu einer Vergrößerung des Vortest-Effekts beitragen können. Des Weiteren ergab sich durch die Befunde der Experimente 5-7 noch die zusätzliche Gemeinsamkeit zwischen dem klassischen Test-Effekt und dem Vortest-Effekt, dass in beiden Paradigmen semantische Mediatoren einerseits in einem Rekognitionstest häufiger falsch als gelernt eingeschätzt werden und andererseits als Hinweisreize den Abruf der Target-Items in den Bedingungen mit Abrufübung erleichtern und so ebenfalls zu den positiven Gedächtniseffekten führen können (z. B. Carpenter, 2011). Auch zur Literatur des vorwärtsgerichteten Test-Effekts liefern die Befunde der Experimente 1-9 eine interessante Parallele. So führt die Bearbeitung eines Tests nach dem Lernen der vorherigen und vor dem Lernen einer nachfolgenden Liste im Test der zuletzt gelernten Liste typischerweise zu weniger Interlist-Intrusionen, d. h. zu weniger falschen Nennungen von Items aus anderen Listen als das wiederholte Lernen der Listen (Ahn & Chan, 2022; Yang et al., 2018). Ebenso führte auch die Bearbeitung von Vortests in den vorliegenden Experimenten zu einer geringeren Anzahl an Intrusionen als das reine Lernen des Materials. Zusammenfassend lässt sich festhalten, dass die Ergebnisse der durchgeführten Experimente eine Replikation der bisherigen Literatur zum Vortest-Effekt darstellen, die bisherigen Befunde durch unterschiedliche Faktoren erweitern und auch zusätzliche Parallelen des Vortest-Paradigmas zum klassischen Test-Effekt und vorwärtsgerichteten Test-Effekt offenbaren konnten.

10.4. Praktische Implikationen

Die Ergebnisse der durchgeführten Experimente verdeutlichen, dass Vortests ein großes Potenzial für den effektiven Einsatz als Lernstrategie in unterschiedlichen Kontexten bergen. Während die bisherige Literatur bereits zeigen konnte, dass Vortests vor Phänomenen wie zeitabhängigem Vergessen und Vergessen durch retroaktive Interferenz schützen können (z. B. Kliegl et al., 2023, 2024), legen die gefundenen Ergebnisse nahe, dass Vortests auch gegenüber Lerndefiziten in Folge von Ablenkung durch eine weitere Aufgabe einen protektiven Faktor darstellen können. Dieser Befund hat insbesondere vor dem Hintergrund, dass in realen Lernumgebungen der Wissenserwerb häufig unter Ablenkung stattfindet (z. B. Dontre, 2021; Wood et al., 2018), eine große Relevanz für die praktische Anwendung dieses

Paradigmas. Allerdings scheint hierbei insbesondere der Zeitpunkt der Ablenkung entscheidend zu sein, da sich dieser Vorteil nur dann zeigte, wenn die Ablenkung während des Vortests, nicht aber, wenn diese während des nachfolgenden Feedbacks stattfand. Dies könnte für Lernende und Lehrpersonen also bedeuten, dass bei dem Einsatz dieser Lerntechnik darauf geachtet werden sollte, dass insbesondere bei der Präsentation der Lerninhalte die volle Aufmerksamkeit für die Enkodierung des Materials zur Verfügung steht, wohingegen eine Ablenkung während des Vortests nicht so schwer ins Gewicht zu fallen scheint. So wäre es beispielsweise denkbar, Vortests gezielt zu Beginn einer Unterrichtsstunde einzusetzen, da hier häufig noch Ablenkung durch Gespräche oder das verspätete Eintreffen einzelner Lernender stattfindet und die direkte Präsentation der Lerninhalte in dieser Phase geteilter Aufmerksamkeit die spätere Erinnerungsleistung beeinträchtigen könnte. Wird hingegen diese Phase durch einen Vortest genutzt, um das Interesse der Lernenden zu wecken, und werden die Lerninhalte erst nachfolgend präsentiert, wenn sich die erste Unruhe gelegt hat, könnte dies sogar zu einem späteren Gedächtnisvorteil führen.

Des Weiteren ergibt sich aus den Ergebnissen die Empfehlung, Vortests beim Lernen nicht nur einmalig anzuwenden, sondern wiederholte Testdurchgänge vor der Präsentation des Lernmaterials zu absolvieren. So könnten Lehrpersonen Lernende beispielsweise dazu auffordern, wiederholt einen Rateversuch für die potenzielle Lösung abzugeben, da dies den positiven Effekt auf die spätere Erinnerungsleistung noch weiter zu verstärken scheint. Insbesondere die Erkenntnis, dass die zusätzlichen Rateversuche nicht zwingend mit einer längeren Vortest- bzw. Lerndauer einhergehen müssen, um diesen Vorteil zu ermöglichen, scheint dabei die Umsetzung und Integration wiederholter Vortests in den Lernalltag deutlich zu erleichtern. Ebenso scheinen Vortests auch zu positiven Transfer-Effekten führen zu können, indem sich die gelernte Zielinformation auch dann abrufen lässt, wenn statt des ursprünglich gelernten Cues ein semantischer Mediator präsentiert wird. Dieser Effekt lässt sich darauf zurückführen, dass im Vortest semantisch verwandte Information aktiviert und in die Cue-Target-Repräsentation integriert wird. Lehrpersonen könnten diesen Transfereffekt didaktisch nutzen, indem sie zu Beginn einer Unterrichtseinheit beispielsweise fragen „Was macht eine Rose aus?“, wodurch Konzepte wie *Pflanze*, *Natur* oder *Wachstum* aktiviert werden könnten. Wenn später im Unterricht das Prinzip der

Photosynthese erklärt und abschließend die Testfrage „Wie gewinnen Tulpen Energie?“ gestellt wird, könnte die Präsentation dieser semantisch mit der Ursprungsfrage verwandten Information die korrekte Antwort („Sie betreibt Photosynthese“) erleichtern. Ohne die Bearbeitung eines Vortests könnte dieser Transfer hingegen deutlich schwerer fallen.

Insgesamt legen die gezeigten Ergebnisse also nahe, dass Vortests sowohl im praktischen Bildungskontext als auch als individuelle Lernstrategie vielfältig eingesetzt werden können und ebenso wie andere Arten der Abrufübung die spätere Erinnerungsleistung positiv beeinflussen können. Daher wäre es wünschenswert, wenn bei einer erneuten Befragung unter Studierenden oder Schülern zur bevorzugten Lernstrategie, die Nutzung von Tests (insbesondere auch von Vortests) deutlich häufiger als bevorzugte und angewandte Lerntechnik genannt werden würde, als dies in der Umfrage von Roediger et al. (2009) der Fall war. Um dies zu gewährleisten, scheinen einige Punkte bei der praktischen Umsetzung des Vortest-Paradigmas von besonderer Relevanz.

So erscheint es sinnvoll, die Lernenden bei der Vorstellung der Lernmethode darauf hinzuweisen, dass diese bei der Bearbeitung des Vortests typischerweise zu einer Vielzahl an Fehlern führt, die allerdings positive Effekte auf die spätere Erinnerungsleistung haben. So könnte die grundsätzliche Tendenz, Fehler beim Lernen vermeiden zu wollen (z. B. Pan et al., 2020), reduziert und das Bewusstsein für die Vorteile dieser Fehler für den Lernerfolg gesteigert werden (Pan & Rivers, 2023), was zu einer häufigeren Anwendung der Technik beitragen könnte. Zudem sollte darauf geachtet werden, dass alle Lernenden einen eigenen Rateversuch generieren, da bisherige Studien darauf hinweisen, dass Fehler insbesondere dann die spätere Erinnerungsleistung fördern, wenn diese selbst generiert werden. Metcalfe und Xu (2018) konnten beispielsweise zeigen, dass Items, für die selbst eine fehlerhafte Antwort gegeben wurde, in einem späteren Test besser erinnert werden konnten als Items, für die der Fehler nur bei einer anderen Versuchsperson beobachtet wurde.

Zusätzlich sollten Lehrpersonen ankündigen, dass Vortests regelmäßig bereits vor dem Lernen der neuen Themen durchgeführt werden. Denn eine solche Ankündigung kann laut Bieleke et al. (2022) die Angst vor Tests reduzieren und so den Nutzen von Vortests, insbesondere im Vergleich zu unerwarteten Tests, weiter steigern. Abschließend gilt es zu beachten, dass die Durchführung von Vortests einen zusätzlichen zeitlichen Aufwand im Vergleich zum unmittelbaren Lernen darstellt. In experimentellen Untersuchungen, die den

vorteilhaften Effekt von Vortests gegenüber einer reinen Lern-Bedingung zeigen, wird die Triallänge dieser beiden Bedingungen typischerweise durch eine Verkürzung der Lernzeit in der Vortest-Bedingung konstant gehalten. Die Ergebnisse von Studien zum Vortest-Effekt legen also nahe, dass Lehrpersonen die zusätzlich erforderliche Zeit für Vortests mit einplanen und durch eine Reduktion der Präsentationsdauer der Lerninhalte ausgleichen könnten, sodass in Summe die Gesamtdauer einer Lerneinheit identisch bleibt. Dies würde eine effektive und umsetzbare Integration von Vortests in den Lehralltag ermöglichen.

10.5. Ansätze für zukünftige Forschung

Die hier berichteten Experimente 1-9 konnten wichtige Erkenntnisse über mögliche zugrundeliegende Mechanismen des Vortest-Effekts liefern. Allerdings ergeben sich aus den gezeigten Befunden auch neue interessante Fragestellungen und Ansätze für zukünftige Forschungen, die das Verständnis für diesen Gedächtniseffekt noch weiter fördern könnten.

Bei der Betrachtung des Vortest-Effekts wird häufig vor allem die Lernleistung der im Vortest präsentierten Inhalte untersucht, wobei die Inhalte der Vortest-Bedingung typischerweise besser abgerufen werden können als die Inhalte der Lern-Bedingung. Es ist jedoch insbesondere für die effiziente Anwendung dieser Lernstrategie in der Praxis von großem Interesse, wie sich die Bearbeitung eines Vortests auch auf ungetestetes Material auswirkt, da sich häufig aus Zeitgründen nicht alle für einen späteren Test relevanten Inhalte vortesten lassen. Sollten sich die positiven Effekte von Vortests auch auf dieses Material transferieren lassen, indem es durch Vortests möglicherweise zu einer Aktivierung und Bildung von Wissensstrukturen und semantischen Netzwerken kommt, in die nachfolgend auch die ungetestete Information integriert und somit besser enkodiert werden könnte, würde dies die praktische Anwendbarkeit dieser Strategie noch weiter steigern. Ein negativer Einfluss von Vortests auf die Lernleistung von ungetesteten Inhalten könnte hingegen bedeuten, dass die Strategie nur gezielt für spezifische Inhalte zum Einsatz kommen sollte. Dies könnte beispielsweise dadurch zustande kommen, dass die Bearbeitung eines Vortests zu einer verbesserten Enkodierung der getesteten Inhalte führt, indem die verfügbaren Aufmerksamkeitsressourcen selektiv auf die Verarbeitung der Inhalte des Vortests gelenkt

werden. Somit könnten für die Enkodierung der ungetesteten Informationen weniger Ressourcen verfügbar und die Lernleistung beeinträchtigt sein (James & Storm, 2019).

Die bisherige Befundlage der Untersuchung von Transfereffekten des Vortest-Effekts zeigt ein inkonsistentes Bild, wobei in einigen Studien ein positiver Transfer auf ungetestetes Material beobachtet werden konnte (z. B. Richland et al., 2009; St. Hilaire et al., 2019). Andere Studien hingegen zeigten keine Transfereffekte (z. B. James & Storm, 2019; Rothkopf, 1966; St. Hilaire & Carpenter, 2020), und in dem vorliegenden Experiment 4 ergab sich sogar eine schlechtere Erinnerungsleistung für ungetestetes Material in der Vortest-Bedingung verglichen mit der Lern-Bedingung.

In einer neuen Studie untersuchten Sana und Carpenter (2023), inwiefern der Zeitpunkt, wann das ungetestete Material in der Lernphase präsentiert wird, eine mögliche moderierende Variable für den Transfer des Vortest-Effekts auf ungetestetes Material darstellt. In zwei Experimenten sollten die Versuchspersonen entweder Text-Passagen oder den Inhalt von Video-Sequenzen lernen, wobei die Hälfte der relevanten Inhalte in einem Vortest abgefragt wurden und die andere Hälfte der Lerninhalte nur in der nachfolgenden Lernphase präsentiert wurden. Dabei wurde variiert, ob die Antwort der Vortest-Fragen unmittelbar zu Beginn der Video- bzw. Textsequenz und somit vor dem ungetesteten Material präsentiert wurden oder am Ende und somit, nachdem die ungetesteten Inhalte zu sehen waren. Die Ergebnisse beider Experimente ergaben eine bessere Erinnerungsleistung für das vorgetestete Material verglichen mit dem ungetesteten Material, wenn dieses zu Beginn der Lernphase präsentiert wurde, und eine vergleichbare Erinnerungsleistung für beide Materialtypen, wenn die Lösung der Vortest-Fragen erst am Ende der Lernphase zu sehen war. Dies wurde von den Autoren vor dem Hintergrund der Aufmerksamkeitsfenster-Hypothese dahingehend interpretiert, dass es durch den Vortest zu einer Öffnung des Aufmerksamkeitsfensters kam, durch das alle nachfolgend präsentierten Inhalte aufmerksamer enkodiert werden, bis das Fenster sich nach der Präsentation der korrekten Lösung wieder schließt. Wird die ungetestete Information also nach dem Vortest und vor dem korrektiven Feedback präsentiert, wird diese ähnlich aufmerksam enkodiert wie die getestete Information, was sich in vergleichbaren Erinnerungsquoten in einem finalen Test niederschlägt (Sana & Carpenter, 2023).

Allerdings wurde in dieser Studie keine reine Lern-Bedingung verwendet, weshalb unklar ist, inwiefern für beide Arten des Lernmaterials tatsächlich ein Vortest-Effekt in Form einer besseren Erinnerungsleistung in der Vortest-Bedingung verglichen mit einer Lern-Bedingung besteht. Hier könnte eine Replikation der Studie von Sana und Carpenter (2023) unter Hinzunahme einer Lern-Bedingung Aufschluss geben.

In bisherigen Studien konnte bereits die Robustheit des Vortest-Effekts für unterschiedliche Test-Formate wie Multiple-Choice-Fragen (z. B. Potts & Shanks, 2014), Cued-Recall-Tests (z. B. Huelser & Metcalfe, 2012; Kliegl et al., 2022), Rekognitionstests (Seabrooke et al., 2021), oder offene Fragen im Kurzantwortformat (Richland et al., 2009) demonstriert werden. Ein direkter Vergleich der Effektivität der unterschiedlichen Test-Formate für den Vortest-Effekt innerhalb eines Experiments wurde bisher jedoch noch nicht durchgeführt. In einer Studie von Seabrooke et al. (2021) wurde zwar sowohl ein Rekognitionstest als auch ein Cued-Recall-Test für den finalen Abruf der Lerninhalte verwendet, allerdings wurde in dieser Studie die Wirksamkeit der beiden Testarten nur getrennt für verwandtes und unverwandtes Lernmaterial untersucht. Dabei zeigte sich im Cued-Recall-Test ausschließlich für verwandtes Material ein Vortest-Effekt, während sich im Rekognitionstest für beide Itemarten ein Vorteil in der Vortest-Bedingung ergab.

Der initiale Vortest wurde in dieser Studie allerdings für alle Items im Cued-Recall-Format durchgeführt, sodass nur für einen Teil der Items das finale Test-Format dem Vortest-Format entsprach, während sich die Formate für die übrigen Items voneinander unterschieden. Darüber hinaus lag die Itemanzahl pro Bedingung aufgrund der Unterteilung in Itemtyp und Testart mit nur sechs Items deutlich unter der typischerweise verwendeten Anzahl an Wortpaaren in Vortest-Studien. Insgesamt lassen sich aus den Befunden dieser Studie folglich nur bedingt Rückschlüsse auf die generelle Wirksamkeit der Testformate für den Vortest-Effekt und darauf, ob die Art der Abrufübung die Größe des Vortest-Effekts beeinflusst, ziehen. Eine Replikation der Studie von Seabrooke et al. (2021) mit einer zusätzlichen Vortest-Bedingung, in der die Versuchspersonen bei jedem Item einschätzen sollen, ob sie dieses Item nachfolgend lernen werden (ähnlich einem Rekognitions-Format), sowie einer höheren Itemanzahl, könnte hier erste Hinweise liefern. Zusätzlich wäre auch die Ergänzung des Studiendesigns um weitere Abrufarten, wie einen Test mit Kurzantworten

oder einen Multiple-Choice-Test, denkbar. Auf diese Weise ließen sich wichtige Erkenntnisse über den optimalen Einsatz von Vortests im pädagogischen Kontext gewinnen.

Einen weiteren interessanten Aspekt für zukünftige Forschung zum Vortest-Paradigma stellt die Darbietungsform von Vortests dar. In den meisten bisherigen Studien wurde der initiale Vortest visuell präsentiert (z. B. Hays et al., 2013), wobei die Fragen oder Items für die Lernenden für mehrere Sekunden sichtbar blieben. Im pädagogischen Alltag werden Lernende jedoch häufig mit mündlichen Fragen konfrontiert und auch Lerninhalte werden nur teilweise visuell, vorrangig aber mündlich dargeboten. Daher wäre es für die Anwendung von Vortests im pädagogischen Lernalltag von großer Relevanz zu untersuchen, ob Lernende auch dann noch von Vortests profitieren, wenn sowohl der Vortest als auch die Präsentation des Lernmaterials mündlich erfolgt. Darüber hinaus stellt sich die Frage, ob die Wirksamkeit von Vortests davon abhängt, ob diese von einer Lehrperson oder einem Versuchsleiter unmittelbar instruiert werden oder ob diese beispielsweise über eine digitale Lern-App stattfinden. Ein direkter Vergleich der Größe des Vortest-Effekts zwischen Items, für die zu Beginn der Lerneinheit per App selbstständig ein Vortest bearbeitet wird und Items, für die der Vortest durch eine Lehrperson oder einen Versuchsleiter bereitgestellt wird, könnte hierfür erste Erkenntnisse liefern.

Nachdem in der vorliegenden Arbeit vor allem die Rolle von Elaborations- und Aufmerksamkeitsprozessen direkt untersucht werden konnte, wäre zukünftig die experimentelle Untersuchung weiterer Erklärungsansätze, wie beispielsweise der Search-Set-Hypothese des Vortest-Effekts, von großem Interesse. Diese Hypothese nimmt an, dass es während des Vortests zur Aktivierung eines Search-Sets kommt, das auch die korrekte Lösung beinhaltet (Grimaldi & Karpicke, 2012). Dieser Prozess ließe sich möglicherweise durch die Erfassung von Reaktionszeiten während der Lernphase experimentell untersuchen.

In einer Vortest-Bedingung könnte den Versuchspersonen zunächst ein Cue-Wort präsentiert werden, wobei für das zugehörige Target-Wort ein Rateversuch abgegeben werden soll (Baum - ?). Anschließend wird das Cue-Item zusammen mit einem verwandten (Baum – Ast) und einem unverwandten Target-Wort (Baum – Hose) präsentiert, wobei die Versuchspersonen für die Hälfte der Wortpaare per Tastendruck so schnell wie möglich das verwandte Wortpaar und für die andere Hälfte der Items das unverwandte Wortpaar auswählen sollen. In einer Lern-Bedingung wird den Versuchspersonen direkt die

Auswahlaufgabe präsentiert. Nachfolgend bleibt das korrekte Wortpaar für einige Sekunden auf dem Bildschirm bestehen, bevor das nächste Item folgt. Abschließend findet ein finaler Test im Cued-Recall-Format statt.

Auf Grundlage der Search-Set-Hypothese wäre zu erwarten, dass es in der Vortest-Bedingung bei verwandten Wortpaaren zu kürzeren Reaktionszeiten kommt als in der Lern-Bedingung, da sich durch die Voraktivierung der korrekten Lösung ein positiver Priming-Effekt (z. B. Neely, 1977) zeigen sollte. Für unverwandte Wortpaare hingegen sollte es zu einer Interferenz der voraktivierten Inhalte mit der korrekten Lösung und einer Hemmung dieser Inhalte durch kognitive Kontrollprozesse kommen. Diese Mechanismen sollten in der Vortest-Bedingung zu einem negativen Priming-Effekt mit langsameren Reaktionszeiten verglichen mit der Lern-Bedingung für diese Items führen. Für die spätere Erinnerungsleistung wäre für verwandte Wortpaare ein Vortest-Effekt, für unverwandte Wortpaare ein Ausbleiben dieses Erinnerungsvorteils in der Vortest-Bedingung zu erwarten, denn gemäß der Search-Set-Hypothese sollten sich die Vorteile durch Vortests nur dann zeigen, wenn das Search-Set die korrekte Target-Information beinhaltet. Dies sollte nur bei zusammenhängenden Wortpaaren der Fall sein.

Ein weiterer Forschungsansatz, um ein besseres Verständnis für die theoretischen Hintergründe des Vortest-Effekts gewinnen zu können, könnte die Untersuchung neuronaler Korrelate, beispielsweise mittels einer funktionellen Magnetresonanztomographie (fMRT), sein. So ließe sich damit untersuchen, welche Gehirnregionen bei der Bearbeitung eines Vortests und der Präsentation des korrektiven Feedbacks in der Vortest-Bedingung stärker aktiv sind als in der reinen Lern-Bedingung. Würde es sich dabei beispielsweise um den anterioren cingulären Kortex (ACC) oder den dorsolateralen medialen Präfrontalkortex handeln, könnte dies ein Indiz dafür sein, dass kognitive Kontrollprozesse eine wichtige Rolle für den Vortest-Effekt spielen, denn diese Regionen sind typischerweise bei der Feststellung eines Fehlers und der nachfolgenden Aktivierung von kognitiven Kontrollprozessen und Fehlerkorrekturmechanismen aktiv (z. B. Blumenfeld & Ranganath, 2007; Botvinick et al., 2001; Yeung et al., 2004).

In einer Studie von Gruber et al. (2014) wurde untersucht, wie subjektiv eingeschätzte Neugierde das Gedächtnis und die spätere Erinnerungsleistung an Wissensfragen beeinflusst und welche Gehirnregionen während dieser Prozesse aktiv sind. Dabei zeigte sich einerseits

eine bessere Erinnerungsleistung für Items, für die die Neugierde als hoch eingeschätzt wurde. Andererseits ergab die Auswertung der fMRT-Daten eine stärkere Aktivierung der Substantia Nigra, des ventralen tegmental Areal und des Nucleus accumbens bei höheren Neugierde-Ratings. Zudem war eine stärkere Aktivierung des Hippocampus und der Verbindungen zum Striatum und in den Präfrontalkortex bei anschließend korrekt abgerufenen Items zu beobachten. Sollte die Bearbeitung eines Vortests, wie von der Aufmerksamkeits-Hypothese angenommen, zu einer gesteigerten Neugier für die korrekte Lösung führen, könnte sich dies also möglicherweise auch in einer stärkeren Aktivierung dieser Gehirnareale in der Vortest-Bedingung niederschlagen, wodurch sich diese Annahme der Aufmerksamkeits-Hypothese zusätzlich überprüfen ließe.

Spielen hingegen Elaborationsprozesse und die semantische Verknüpfung der Informationen eine entscheidende Rolle für die Entstehung des Vortest-Effekts, wäre es auch denkbar, dass Regionen, die mit diesen Prozessen in Verbindung stehen, in der Vortest-Bedingung stärker aktiv sind als in der Lern-Bedingung. Es könnte sich dabei vor allem um Verknüpfungen des lateralen Temporalkortex und des medialen Präfrontalkortex mit dem Hippocampus (Wing et al., 2013; van den Broek et al., 2016) sowie um den Gyrus angularis, den ventro-temporalen Cortex, den Gyrus frontalis inferior und den Gyrus cinguli (Binder et al., 2009) handeln. Denn für diese Regionen konnte in einer Meta-Analyse von Binder et al. (2009), in der 120 fMRT-Studien betrachtet wurden, ein Zusammenhang mit der Aktivierung von semantischen Verarbeitungs- und Integrationsprozessen hergestellt werden. Auch wenn die Aktivierung der einzelnen Gehirnareale nicht ausschließlich einem Prozess zugeordnet werden kann und unterschiedliche Gehirnstrukturen miteinander interagieren, könnten doch hirnpfysiologische Befunde oder eine regionale Eingrenzung der Aktivierung einen weiteren Einblick in die grundlegend relevanten Mechanismen des Vortest-Effekts ermöglichen.

10.6. Fazit

Die Ergebnisse der vorliegenden Experimente konnten zeigen, dass unterschiedliche Faktoren die Effektivität von Vortests beeinflussen können. So ergab sich bei der Betrachtung wiederholt durchgeführter Vortests eine Zunahme des Vortest-Effekts mit steigender Anzahl an Testdurchläufen. Außerdem führte auch die Präsentation semantischer

Mediatoren zu einem verbesserten Abruf der Zielinformationen, was die Annahme stützt, dass es während eines Vortests zur Aktivierung eines semantischen Netzwerks und der Integration der Lerninhalte in dieses Netzwerk kommt. Zudem deuten die gefundenen Ergebnisse darauf hin, dass Vortests vor den negativen Auswirkungen geteilter Aufmerksamkeit auf die spätere Erinnerungsleistung schützen können und eine Störung während des Vortests sogar zu einem noch größeren Vortest-Effekt führen kann. Insbesondere die Enkodierung der Lerninhalte während des korrektiven Feedbacks scheint jedoch Aufmerksamkeitsressourcen zu erfordern, da eine Ablenkung hier zu einem Ausbleiben der vorteilhaften Effekte dieser Abrufübung führt. Insgesamt lassen sich die Befunde der vorliegenden Arbeit also gut mit der Elaborations- und der Aufmerksamkeits-Hypothese des Vortest-Effekts vereinbaren und konnten somit wichtige Hinweise für die noch ungeklärte Frage nach den kognitiven Mechanismen des Vortest-Effekts liefern. Die vorliegenden Experimente konnten zudem die effiziente Anwendbarkeit von Vortests im praktischen Bildungskontext untermauern, sodass es wünschenswert wäre, wenn Lernende auf die eingangs beschriebene Aussage „Unterlagen weg, wir schreiben erstmal einen kurzen Test.“ zukünftig nicht mehr mit einem enttäuschten „Oh nein, muss das sein?“ sondern mit einem erfreuten „Super, davon kann ich profitieren!“ reagieren würden.

Literaturverzeichnis

- Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research*, 87(3), 659-701.
- Agarwal, P. K., Karpicke, J. D., Kang, S. H. K., Roediger, H. L., & McDermott, K. B. (2008). Examining the Test effect with open- and closed-book tests. *Applied Cognitive Psychology*, 22, 861-876.
- Agarwal, P. K., Nunes, L. D., & Blunt, J. R. (2021). Retrieval practice consistently benefits student learning: A systematic review of applied research in schools and classrooms. *Educational Psychology Review*, 33(4), 1409-1453.
- Ahn, D., & Chan, J. C. (2022). Does testing enhance new learning because it insulates against proactive interference? *Memory & Cognition*, 50(8), 1664–1682.
- Allen, G. A., Mahler, W. A., & Estes, W. K. (1969). Effects of recall tests on long-term retention of paired associates. *Journal of Verbal Learning and Verbal Behavior*, 8, 463-470.
- Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, 22(3), 261-295.
- Anderson, N. D., Craik, F. I., & Naveh-Benjamin, M. (1998). The attentional demands of encoding and retrieval in younger and older adults: I. Evidence from divided attention costs. *Psychology and Aging*, 13(3), 405.
- Aslan, A., & Bäuml, K.-H. T. (2016). Testing enhances subsequent learning in older but not in younger elementary school children. *Developmental Science*, 19, 992-998.
- Bäckman, L., Jones, S., Berger, A. K., Laukka, E. J., & Small, B. J. (2004). Multiple cognitive deficits during the transition to Alzheimer's disease. *Journal of internal medicine*, 256(3), 195-204.

- Balota, D. A., Duchek, J. M., Sergent-Marshall, S. D., & Roediger III, H. L. (2006). Does expanded retrieval produce benefits over equal-interval spacing? Explorations of spacing effects in healthy aging and early stage Alzheimer's disease. *Psychology and Aging, 21*(1), 19-31.
- Balota, D. A., & Lorch, R. F. (1986). Depth of automatic spreading activation: Mediated priming effects in pronunciation but not in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*(3), 336-345.
- Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods, 39*(3), 445-459.
- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice, 18*(1), 5-25.
- Bieleke, M., Schwarzkopf, J. M., Götz, T., & Haag, L. (2022). The agonizing effects of uncertainty: Effects of announced vs. unannounced performance assessments on emotions and achievement. *Plos one, 17*(8), e0272443.
- Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex, 19*(12), 2767-2796.
- Bjork, R. A. (1975). Retrieval as a memory modifier. In R. Solso (Hrsg.), *Information processing and cognition: The Loyola Symposium* (S. 123-144). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Bjork, R. A. (1988). Retrieval practice and the maintenance of knowledge. *Practical aspects of memory: Current research and issues, 1*, 396-401.
- Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Hrsg.), *Metacognition: Knowing about knowing* (S. 185-205). Cambridge, MA: MIT Press.

- Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher & J. R. Pomerantz (Hrsg.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (S.56-64). New York: Worth.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74.
- Blumenfeld, R. S., & Ranganath, C. (2007). Prefrontal cortex and long-term memory encoding: an integrative review of findings from neuropsychology and neuroimaging. *The Neuroscientist*, 13(3), 280-291.
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, 108(3), 624-652.
- Boustani, S., & Shanks, D. R. (2022). Heterogeneity and publication bias in research on test-potentiated new learning. *Collabra: Psychology*, 8(1), 31996.
- Bridger, E. K., & Mecklinger, A. (2014). Errorful and errorless learning: The impact of cue-target constraint in learning from errors. *Memory & Cognition*, 42(6), 898-911.
- Buchin, Z. L., & Mulligan, N. W. (2017). The testing effect under divided attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(12), 1934-1947.
- Buchin, Z. L., & Mulligan, N. W. (2019). Divided attention and the encoding effects of retrieval. *Quarterly Journal of Experimental Psychology*, 72(10), 2474-2494.
- Carneiro, P., Lapa, A., & Finn, B. (2018). The effect of unsuccessful retrieval on children's subsequent learning. *Journal of Experimental Child Psychology*, 166, 400-420.
- Carpenter, S. K. (2009). Cue strength as a moderator of the Test effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(6), 1563-1569.
- Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the Test effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(6), 1547-1552.

- Carpenter, S. K., Pashler, H., Wixted, J. T., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory & Cognition*, 36(2), 438-448.
- Carpenter, S. K., Rahman, S., & Perkins, K. (2018). The effects of prequestions on classroom learning. *Journal of Experimental Psychology: Applied*, 24(1), 34-42.
- Carpenter, S. K., & Toftness, A. R. (2017). The effect of prequestions on learning from video presentations. *Journal of Applied Research in Memory and Cognition*, 6(1), 104-109.
- Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition*, 20, 633-642.
- Chan, J. C., Meissner, C. A., & Davis, S. D. (2018b). Retrieval potentiates new learning: A theoretical and meta-analytic review. *Psychological Bulletin*, 144, 1111-1146.
- Clark, C. M. (2016). *When and Why does Learning Profit from the Introduction of Errors?* University of California, Los Angeles.
- Collins, A. M., & Quillian, M. R. (1972). Experiments on semantic memory and language comprehension. In L. Gregg (Hrsg.), *Cognition and Learning* (S. 117-138). New York, NY: Wiley.
- Craik, F. M., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). Encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General*, 125(2), 159-180.
- Cyr, A. A., & Anderson, N. D. (2012). Trial-and-error learning improves source memory among young and older adults. *Psychology and Aging*, 27(2), 429-439.
- Cyr, A. A., & Anderson, N. D. (2015). Mistakes as stepping stones: Effects of errors on episodic memory among younger and older adults. *Journal of Experimental Psychology: Learning Memory and Cognition*, 41(3), 841-850.
- Cyr, A. A., & Anderson, N. D. (2018). Learning from your mistakes: Does it matter if you're out in left foot, I mean field? *Memory*, 26(9), 1281-1290.
- Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227-268.

- de Lima, N. K., & Jaeger, A. (2020). The effects of prequestions versus postquestions on memory retention in children. *Journal of Applied Research in Memory and Cognition*, 9(4), 555-563.
- de Jonge, M., & Tabbers, H. K. (2013). Repeated testing, item selection, and relearning. *Experimental Psychology*, 60(3), 206-212.
- DiMarco, D., & Marmurek, H. (2022). Transfer of test-enhanced learning: Limitations of the mediator effectiveness hypothesis. *Experimental Psychology*, 69(3), 119-131.
- Dontre, A. J. (2021). The influence of technology on academic distraction: A review. *Human Behavior and Emerging Technologies*, 3(3), 379-390.
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14(1), 4-58.
- Ebbinghaus, H. (1885). *Über das Gedächtnis*. Leipzig: Duncker & Humblot.
- Endres, T., Carpenter, S., Martin, A., & Renkl, A. (2017). Enhancing learning by retrieval: Enriching free recall with elaborative prompting. *Learning and Instruction*, 49, 13-20.
- Estes, W. K. (1955). Statistical theory of spontaneous recovery and regression. *Psychological Review*, 62, 145-154.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149-1160.
- Fernandes, M. A., & Moscovitch, M. (2003). Interference effects from divided attention during retrieval in younger and older adults. *Psychology and Aging*, 18(2), 219-230.
- Frase, L. T. (1968). Effect of question location, pacing, and mode upon retention of prose material. *Journal of Educational Psychology*, 59(4), 244-249.
- Grimaldi, P. J., & Karpicke, J. D. (2012). When and why do retrieval attempts enhance subsequent encoding? *Memory & Cognition*, 40, 505-513.

- Gruber, M. J., Gelman, B. D., & Ranganath, C. (2014). States of curiosity modulate hippocampus-dependent learning via the dopaminergic circuit. *Neuron*, 84(2), 486-496.
- Guthrie, E. (1952). *The psychology of learning* (Hrsg.). New York: Harper.
- Halamish, V., & Bjork, R. A. (2011). When does testing enhance retention? A distribution-based interpretation of retrieval as a memory modifier. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(4), 801-812.
- Hartley, J. (1973). The effect of pre-testing on post-test performance. *Instructional Science*, 2(2), 193-214.
- Hausman, H., & Rhodes, M. G. (2018). When pretesting fails to enhance learning concepts from reading texts. *Journal of Experimental Psychology: Applied*, 24(3), 331-346.
- Hays, M. J., Kornell, N., & Bjork, R. A. (2013). When and why a failed test potentiates the effectiveness of subsequent study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(1), 290-296.
- Heritage, M. (2007). Formative assessment: What do teachers need to know and do? *Phi Delta Kappan*, 89(2), 140-145.
- Hester, R., Foxe, J. J., Molholm, S., Shpaner, M., & Garavan, H. (2005). Neural mechanisms involved in error processing: a comparison of errors made with and without awareness. *Neuroimage*, 27(3), 602-608.
- Hopper, W. J., & Huber, D. E. (2018). Learning to recall: Examining recall latencies to test an intra-item learning theory of Test effects. *Journal of Memory and Language*, 102, 1-15.
- Howieson, D. B., Dame, A., Camicioli, R., Sexton, G., Payami, H., & Kaye, J. A. (1997). Cognitive markers preceding Alzheimer's dementia in the healthy oldest old. *Journal of the American Geriatrics Society*, 45(5), 584-589.
- Huelser, B. J., & Metcalfe, J. (2012). Making related errors facilitates learning, but learners do not know it. *Memory & Cognition*, 40(4), 514-527.

- Hunt, R. R., Smith, R. E., & Dunlap, K. R. (2011). How does distinctive processing reduce false recall? *Journal of Memory and Language*, 65(4), 378-389.
- Internationale Hochschule (2022). *Prüfungsangst. Die Fakten. Kurzstudie.* <https://static.iu.de/studies/Pruefungsangst.pdf>. Zugriffen: 29. April 2025
- James, K. K. (2022). *Do More Guesses Lead to More Benefits? On the Consequences of Multiple Attempts in the Pretesting Effect.* University of California, Santa Cruz.
- James, K. K., & Storm, B. C. (2019). Beyond the pretesting effect: What happens to the information that is not pretested? *Journal of Experimental Psychology: Applied*, 25(4), 576-587.
- Jang, Y., & Huber, D. E. (2008). Context retrieval and context change in free recall: Recalling from long-term memory drives list isolation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(1), 112-127.
- Johnson, C. I., & Mayer, R. E. (2009). A testing effect with multimedia learning. *Journal of Educational Psychology*, 101(3), 621-629.
- Karpicke, J. D. (2017). Retrieval-based learning: A decade of progress. In J. H. Byrne (Hrsg.), *Cognitive psychology of memory, Vol. 2 of Learning and memory: A comprehensive reference* (S. 1-26). Elsevier.
- Karpicke, J. D., Butler, A. C., & Roediger III, H. L. (2009). Metacognitive strategies in student learning: do students practise retrieval when they study on their own? *Memory*, 17(4), 471-479.
- Karpicke, J. D., & Roediger III, H. L. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, 57(2), 151-162.
- Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, 319, 966-968.
- Kirk-Johnson, A., Galla, B. M., & Fraundorf, S. H. (2019). Perceiving effort as poor learning: The misinterpreted-effort hypothesis of how experienced effort and perceived learning relate to study strategy choice. *Cognitive Psychology*, 115, 101237.

- Kliegl, O., Bartl, J., & Bäuml, K.-H.T. (2023). The pretesting effect thrives in the presence of competing information. *Memory*, 31(5), 705-714.
- Kliegl, O., Bartl, J., & Bäuml, K.-H.T. (2024). The pretesting effect comes to full fruition after prolonged retention interval. *Journal of Applied Research in Memory and Cognition*, 13(1), 63-70.
- Knight, J. B., Ball, B. H., Brewer, G. A., DeWitt, M. R., & Marsh, R. L. (2012). Testing unsuccessfully: A specification of the underlying mechanisms supporting its influence on retention. *Journal of Memory and Language*, 66(4), 731-746.
- Kornell, N. (2014). Attempting to answer a meaningful question enhances subsequent learning even when feedback is delayed. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(1), 106-114
- Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(4), 989-998.
- Kornell, N., & Vaughn, K. E. (2016). How retrieval attempts affect learning: A review and synthesis. *Psychology of Learning and Motivation*, 65, 183-215.
- Larsen, D. P., Butler, A. C., Lawson, A. L., & Roediger, H. L. (2013). The importance of seeing the patient: test-enhanced learning with standardized patients and written tests improves clinical application of knowledge. *Advances in Health Sciences Education*, 18, 409-425.
- Levine, B., Cabeza, R., McIntosh, A. R., Black, S. E., Grady, C. L., & Stuss, D. T. (2002). Functional reorganisation of memory after traumatic brain injury: a study with H2150 positron emission tomography. *Journal of Neurology, Neurosurgery & Psychiatry*, 73(2), 173-181.
- Macmillan, N. A., & Creelman, C. D. (2021). *Detection theory: A user's guide* (2. Edition). Lawrence Erlbaum Inc.
- Marsh, E. J., Roediger, H. L., III, Bjork, R. A., & Bjork, E. L. (2007). Memorial consequences of multiple-choice testing. *Psychonomic Bulletin & Review*, 14, 194-199.

- McClelland, J. L., & Rumelhart, D. E. (1986). A distributed model of human learning and memory. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group (Hrsg.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 2. Psychological and biological models* (S. 170-215). Cambridge, MA: MIT Press.
- McDaniel, M. A., Kowitz, M. D., & Dunay, P. K. (1989). Altering memory through recall: The effects of cue-guided retrieval processing. *Memory & Cognition*, 17(4), 423-434.
- McDaniel, M. A., Howard, D. C., & Einstein, G. O. (2009). The read-recite-review study strategy: Effective and portable. *Psychological Science*, 20(4), 516-522.
- Mensink, G. J. M., & Raaijmakers, J. G. W. (1988). A model for interference and forgetting. *Psychological Review*, 95(4), 434-455.
- Mera, Y., Rodríguez, G., & Marin-Garcia, E. (2022). Unraveling the benefits of experiencing errors during learning: Definition, modulating factors, and explanatory theories. *Psychonomic Bulletin & Review*, 29(3), 753-765.
- Metcalf, J., & Finn, B. (2012). Hypercorrection of high confidence errors in children. *Learning and Instruction*, 22(4), 253-261.
- Metcalf, J., & Huelser, B. J. (2020). Learning from errors is attributable to episodic recollection rather than semantic mediation. *Neuropsychologia*, 138, 107296.
- Metcalf, J., & Kornell, N. (2007). Principles of cognitive science in education: The effects of generation, errors, and feedback. *Psychonomic Bulletin & Review*, 14, 225-229.
- Metcalf, J., & Xu, J. (2018). Learning from one's own errors and those of others. *Psychonomic Bulletin & Review*, 25(1), 402-408.
- Meyer, A. N., & Logan, J. M. (2013). Taking the testing effect beyond the college freshman: benefits for lifelong learning. *Psychology and Aging*, 28(1), 142-147.
- Middleton, E. L., Schwartz, M. F., Rawson, K. A., Traut, H., & Verkuilen, J. (2016). Towards a theory of learning for naming rehabilitation: Retrieval practice and spacing effects. *Journal of Speech, Language, and Hearing Research*, 59(5), 1111-1122.

- Minkley, N., & Kirchner, W. H. (2012). Influence of test tasks with different cognitive demands on salivary cortisol concentrations in school students. *International Journal of Psychophysiology*, 86(3), 245-250.
- Mulligan, N. W. (2008). Attention and memory. In H. L. Roediger (Hrsg.), *Learning and memory: A comprehensive reference* (S. 7-22). Elsevier.
- Mulligan, N. W., & Buchin, Z. L. (2024). Attention and the forward testing effect. *Memory & Cognition*, 1-16.
- Mulligan, N. W., Buchin, Z. L., & West, J. T. (2022a). Attention, the testing effect, and retrieval-induced forgetting: Distraction dissociates the positive and negative effects of retrieval on subsequent memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 48(12), 1905-1922.
- Mulligan, N. W., Buchin, Z. L., & Zhang, A. L. (2022b). The testing effect with free recall: Organization, attention, and order effects. *Journal of Memory and Language*, 125, 104333.
- Mulligan, N. W., & Picklesimer, M. (2016). Attention and the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(6), 938-950.
- Naveh-Benjamin, M., Kilb, A., & Fisher, T. (2006). Concurrent task effects on memory encoding and retrieval: Further support for an asymmetry. *Memory & Cognition*, 34, 90-101.
- Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited capacity attention. *Journal of Experimental Psychology: General*, 106(3), 226-254.
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms*.
- Nelson, T. O., & Dunlosky, J. (1994). Norms of paired-associate recall during multitrial learning of Swahili–English translation equivalents. *Memory*, 2(3), 325-335.

- Overman, A. A., Stephens, J. D., & Bernhardt, M. F. (2021). Enhanced memory for context associated with corrective feedback: Evidence for episodic processes in errorful learning. *Memory*, 29(8), 1017-1042.
- Overoye, A. L., James, K. K., & Storm, B. C. (2021). A little can go a long way: Giving learners some context can enhance the benefits of pretesting. *Memory*, 29(9), 1206-1215.
- Pan, S. C., & Carpenter, S. K. (2023). Prequestioning and pretesting effects: A review of empirical research, theoretical perspectives, and implications for educational practice. *Educational Psychology Review*, 35(4), 97.
- Pan, S. C., & Rivers, M. L. (2023). Metacognitive awareness of the pretesting effect improves with self-regulation support. *Memory & Cognition*, 51(6), 1461-1480.
- Pan, S. C., Sana, F., Samani, J., Cooke, J., & Kim, J. A. (2020). Learning from errors: Students' and instructors' practices, attitudes, and beliefs. *Memory*, 28(9), 1105-1122.
- Pastötter, B., & Bäuml, K. H. T. (2014). Retrieval practice enhances new learning: The forward effect of testing. *Frontiers in Psychology*, 5, 83305.
- Pastötter, B., Weber, J., & Bäuml, K. H. T. (2013). Using testing to improve learning after severe traumatic brain injury. *Neuropsychology*, 27(2), 280-285.
- Perry, R. J., & Hodges, J. R. (1999). Attention and executive deficits in Alzheimer's disease: A critical review. *Brain*, 122(3), 383-404.
- Postman, L., & Underwood, B. J. (1973). Critical issues in interference theory. *Memory & Cognition*, 1(1), 19-40.
- Potts, R., Davies, G., & Shanks, D. R. (2019). The benefit of generating errors during learning: What is the locus of the effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(6), 1023-1041.
- Potts, R., & Shanks, D. R. (2012). Can testing immunize memories against interference? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(6), 1780-1785.

- Potts, R., & Shanks, D. R. (2014). The benefit of generating errors during learning. *Journal of Experimental Psychology: General*, 143(2), 644-667.
- Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language*, 60(4), 437-447.
- Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science*, 330(6002), 335-335.
- Pressley, M., Tanenbaum, R., McDaniel, M. A., & Wood, E. (1990). What happens when university students try to answer prequestions that accompany textbook material? *Contemporary Educational Psychology*, 15, 27-35.
- Richland, L. E., Kornell, N., & Kao, L. S. (2009). The pretesting effect: Do unsuccessful retrieval attempts enhance learning? *Journal of Experimental Psychology: Applied*, 15(3), 243-257.
- Rickards, J. P. (1979). Adjunct postquestions in text: A critical review of methods and processes. *Review of Educational Research*, 49(2), 181-196.
- Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, 15(1), 20-27.
- Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17, 249-255.
- Rogalski, Y., Altmann, L. J., & Rosenbek, J. C. (2014). Retrieval practice and testing improve memory in older adults. *Aphasiology*, 28(4), 381-400.
- Rothkopf, E. Z. (1966). Learning from written instructive materials: An exploration of the control of inspection behavior by test-like events. *American Educational Research Journal*, 3(4), 241-249.
- Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, 140(6), 1432-1463.
- Rubin, D. C., & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review*, 103, 734-760.

- Sana, F., & Carpenter, S. K. (2023). Broader benefits of the pretesting effect: Placement matters. *Psychonomic Bulletin & Review*, 30(5), 1908-1916.
- Seabrooke, T., Mitchell, C. J., Wills, A. J., & Hollins, T. J. (2021). Pretesting boosts recognition, but not cued recall, of targets from unrelated word pairs. *Psychonomic Bulletin & Review*, 28(1), 268-273.
- Seabrooke, T., Mitchell, C. J., Wills, A. J., Waters, J. L., & Hollins, T. J. (2019). Selective effects of errorful generation on recognition memory: The role of motivation and surprise. *Memory*, 27(9), 1250-1262.
- Skinner, B. F. (1958). Teaching machines: From the experimental study of learning come devices which arrange optimal conditions for self-instruction. *Science*, 128, 969-977.
- Slamecka, N. J., & Fevreiski, J. (1983). The generation effect when generation fails. *Journal of Verbal Learning and Verbal Behavior*, 22(2), 153-163.
- Small, B. J., Herlitz, A., Fratiglioni, L., Almkvist, O., & Bäckman, L. (1997). Cognitive predictors of incident Alzheimer's disease: a prospective longitudinal study. *Neuropsychology*, 11(3), 413-420.
- Soderstrom, N. C., & Bjork, E. L. (2023). Pretesting Enhances Learning in the Classroom. *Educational Psychology Review*, 35(3), 88.
- Spitzer, B., & Bäuml, K. H. (2009). Retrieval-induced forgetting in a category recognition task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(1), 286-291.
- St. Hilaire, K. J. S., & Carpenter, S. K. (2020). Prequestions enhance learning, but only when they are remembered. *Journal of Experimental Psychology: Applied*, 26(4), 705-716.
- St. Hilaire, K. J., Carpenter, S. K., & Jennings, J. M. (2019). Using prequestions to enhance learning from reading passages: The roles of question type and structure building ability. *Memory*, 27(9), 1204-1213.
- Sumowski, J. F., Leavitt, V. M., Cohen, A., Paxton, J., Chiaravalloti, N. D., & DeLuca, J. (2013). Retrieval practice is a robust memory aid for memory-impaired patients with MS. *Multiple Sclerosis Journal*, 19(14), 1943-1946.

- Sumowski, J. F., Wood, H. G., Chiaravalloti, N., Wylie, G. R., Lengenfelder, J., & Deluca, J. (2010). Retrieval practice: A simple strategy for improving memory after traumatic brain injury. *Journal of the International Neuropsychological Society*, 16(6), 1147-1150.
- Szpunar, K. K., McDermott, K. B., Roediger, H. L., & III. (2008). Testing during study insulates against the buildup of proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(6), 1392-1399.
- Toftness, A. R., Carpenter, S. K., Lauber, S., & Mickes, L. (2018). The limited effects of prequestions on learning from authentic lecture videos. *Journal of Applied Research in Memory and Cognition*, 7(3), 370-378.
- Unsworth, N., & McMillan, B. D. (2017). Attentional disengagements in educational contexts: A diary investigation of everyday mind-wandering and distraction. *Cognitive research: Principles and implications*, 2(1), 32.
- Van den Broek, G., Takashima, A., Wiklund-Hörnqvist, C., Wirebring, L. K., Segers, E., Verhoeven, L., & Nyberg, L. (2016). Neurocognitive mechanisms of the “testing effect”: A review. *Trends in Neuroscience and Education*, 5(2), 52-66.
- Vaughn, K. E., Hausman, H., & Kornell, N. (2017). Retrieval attempts enhance learning regardless of time spent trying to retrieve. *Memory*, 25(3), 298-316.
- Vaughn, K. E., & Rawson, K. A. (2011). Diagnosing criterion-level effects on memory: What aspects of memory are enhanced by repeated retrieval? *Psychological Science*, 22(9), 1127-1131.
- Vaughn, K. E., & Rawson, K. A. (2012). When is guessing incorrectly better than studying for enhancing memory? *Psychonomic Bulletin & Review*, 19, 899-905.
- Willis, R., Erwin, D., & Adelaja, F. (2020). Struggling prior to a teaching event results in superior short-term skills acquisition in novice learners. *Journal of Surgical Education*, 77(1), 34-39.
- Wing, E. A., Marsh, E. J., & Cabeza, R. (2013). Neural correlates of retrieval-based memory enhancement: an fMRI study of the testing effect. *Neuropsychologia*, 51(12), 2360-2370.

- Wissman, K. T., Rawson, K. A., & Pyc, M. A. (2011). The interim test effect: Testing prior material can facilitate the learning of new material. *Psychonomic Bulletin & Review*, 18, 1140-1147.
- Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*, 114(1), 152-176.
- Wood, E., Mirza, A., & Shaw, L. (2018). Using technology to promote classroom instruction: Assessing incidences of on-task and off-task multitasking and learning. *Journal of Computing in Higher Education*, 30, 553-571.
- Yan, V. X., Rea, S. D., & Sana, F. (2025). The pretesting effect is robust throughout adulthood, but metacognitive beliefs about pretesting and challenge differ. *Learning and Individual Differences*, 120, 102683.
- Yan, V. X., Yu, Y., Garcia, M. A., & Bjork, R. A. (2014). Why does guessing incorrectly enhance, rather than impair, retention? *Memory & Cognition*, 42(8), 1373-1383.
- Yang, C., Luo, L., Vadillo, M. A., Yu, R., & Shanks, D. R. (2021). Testing (quizzing) boosts classroom learning: A systematic and meta-analytic review. *Psychological Bulletin*, 147(4), 399-435.
- Yang, C., Potts, R., & Shanks, D. R. (2017). Metacognitive unawareness of the errorful generation benefit and its effects on self-regulated learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(7), 1073-1092.
- Yang, C., Potts, R., & Shanks, D. R. (2018). Enhancing learning and retrieval of new information: A review of the forward testing effect. *NPJ science of learning*, 3(1), 8.
- Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: conflict monitoring and the error-related negativity. *Psychological Review*, 111(4), 931-959.
- Zawadzka, K., & Hanczakowski, M. (2019). Two routes to memory benefits of guessing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(10), 1748-1760.